

AD-A186 214

IMAGE UNDERSTANDING BY IMAGE-SEEKING ADAPTIVE NETWORKS
(ISAN)(U) MASSACHUSETTS UNIV AMHERST DEPT OF COMPUTER
AND INFORMATION S. D N SPINELLI 10 AUG 87

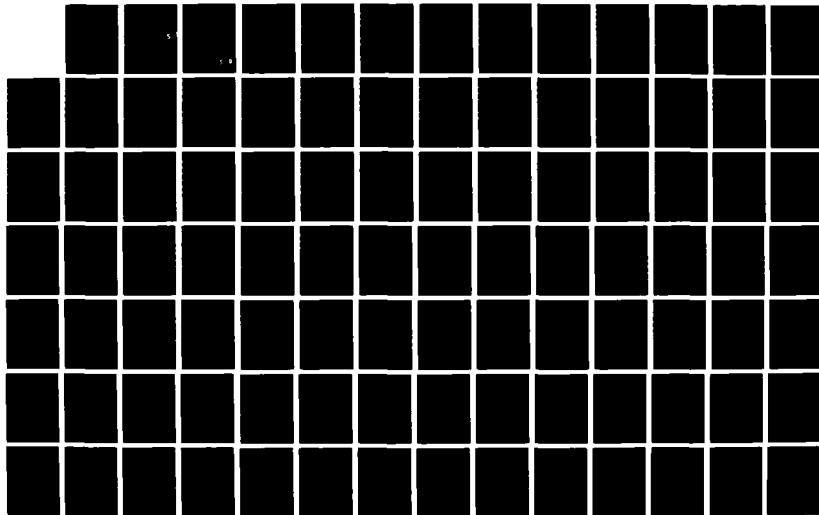
1/2

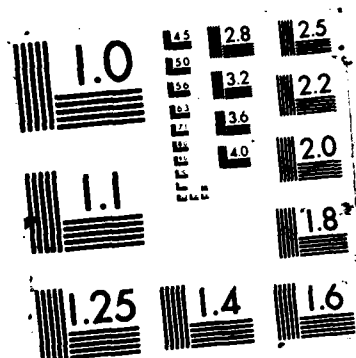
UNCLASSIFIED

AFOSR-TR-87-1286 #AFOSR-83-0207

F/G 17/11

NL





AD-A186 214

FILE COPY

②

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE				
1a. REPORT SECURITY CLASSIFICATION Unclassified		1d. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION AVAILABILITY OF REPORT		
2b. DECLASSIFICATION DOWNGRADING SCHEDULE		Approved for public release; distribution unlimited.		
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 87-1286		
6a. NAME OF PERFORMING ORGANIZATION Computer & Information Science Department, U of MA		6b. OFFICE SYMBOL (If applicable)		7a. NAME OF MONITORING ORGANIZATION AFOSR/NL
6c. ADDRESS (City, State and ZIP Code) University of Massachusetts Amherst, MA 01003		7b. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB, D.C. 20332-6448		
8a. NAME OF FUNDING SPONSORING ORGANIZATION AFOSR/NL		8b. OFFICE SYMBOL (If applicable)		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-83-0207
8c. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB DC 20332-6448		10. SOURCE OF FUNDING NOS.		
11. TITLE (Include Security Classification) Image Understanding by Image Seeking Adaptive Networks ISAN		PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2132	TA N A1
12. PERSONAL AUTHOR(S) D. N. Spinelli				
13a. TYPE OF REPORT Final Technical		13b. TIME COVERED FROM 83MAY01 TO 86APR30		14. DATE OF REPORT (Yr., Mo., Day) 87AUG10
15. PAGE COUNT 116				
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB GR	Image-seeking devices Adaptive networks	
			Image-understanding	
			Differentiated-image amplification	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) A remarkably simple, experimentally inspired, new theory of vision is presented. The theory takes into account the parallel architecture, the adaptive phenomena and the efferent control systems which have been demonstrated in the vision systems of organisms. Further, the complexities of visual receptive fields are made use of to explain the speed, noise resistance, constancies, and holistic aspects of perception. In this theory image understanding is achieved by image-seeking adaptive networks that differentially amplify images of interest without first breaking them down into elementary components. A computer implementation of the theory demonstrates that the mechanisms postulated are feasible. A number of experiments with the model address critical aspects of image understanding and demonstrate that images of interest are captured reliably even in large amounts of noise, or in spite of position and/or size changes. Subjective edges, and other Gestalt-like images, i.e., horizon and terrain, are also seen by ISAN's basic network. Some implications for general vision are outlined.				
20. DISTRIBUTION AVAILABILITY OF ABSTRACT UNCLASSIFIED UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> OTIC USERS <input type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL WILLIAM O. BERRY			22b. TELEPHONE NUMBER (202) 767-5021	22c. OFFICE SYMBOL NL

DD FORM 1473, 83 APR

EDITION OF 1 JAN 73 IS OBSOLETE

Unclassified

19

11 SEP 1987

SECURITY CLASSIFICATION OF THIS PAGE

AFOSR-TR- 87 - 1286

IMAGE UNDERSTANDING BY IMAGE-SEEKING ADAPTIVE NETWORKS ISAN

D. N. Spinelli
Computer and Information Science
University of Massachusetts
Amherst, MA 01003



August 1987

FINAL REPORT for period May 1983 - April 1986

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Approved for public release; distribution unlimited.

LIFE SCIENCES DIRECTORATE
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
BUILDING 410
BOLLING AIR FORCE BASE, D.C. 20332

DTIC
ELECTE
SEP 29 1987
S D
E

87 9 25 026
11 SEP 1987

FINAL REPORT: AFOSR-83-0207

**Image Understanding
by
Image Seeking Adaptive Networks
ISAN**

D. N. Spinelli

Abstract

A remarkably simple, experimentally inspired, new theory of vision is presented. The theory takes into account the parallel architecture, the adaptive phenomena and the efferent control system which have been demonstrated in the vision systems of organisms. Further the complexities of visual receptive fields are made use of to explain the speed, noise resistance, constancies and holistic aspects of perception. In this theory image understanding is achieved by image seeking adaptive networks that differentially amplify images of interest without first breaking them down into elementary components. A computer implementation of the theory demonstrates that the mechanisms postulated are feasible. A number of experiments with the model address critical aspects of image understanding and demonstrate that images of interest are captured reliably even in large amounts of noise, or in spite of position and/or size changes. Subjective edges, and other Gestalt like images, i.e. horizon and terrain are also seen by ISAN's basic network. Some implications for general vision are outlined.

INTRODUCTION

Vision provides the most direct, early and far reaching information as to the three dimensional structure of the environment. Through vision animals are enabled to plan actions necessary to navigate, hunt and/or escape as necessary. These functions, critical for survival, are often performed with astonishing sophistication by the minutest of brains, consider for example the brain of a bee. Vision has emerged no less than forty times during the course of animal evolution-- thus it must be easy to invent and advantageous to possess it. The speed and reliability of animal vision in very many environments, notwithstanding slow and unreliable components, perforce requires massive parallelism, simple circuitry and plasticity.

Understanding animal vision systems is important for many reasons-- from a theoretical point of view it would vastly increase our understanding of how brains store and access vast amounts of complex, multidimensional information; build internal models of the environment, plan and execute visually guided motor sequences, and much more. From the practical point of view knowledge of the mechanisms responsible for image understanding would undoubtedly have important applications in the design of parallel, extremely fast and damage resistant computer

architectures. Finally displays designed to suit the functional and structural architecture of the vision system could substantively increase both the amount of information that can be absorbed by and the performance of the viewer, particularly in terms of reaction time.

In this report we move on two dimensions-- in one we report on some aspects of our reserch on adaptive neural networks in the visual and sensory-motor cortex of cats. **We demonstrate that, under certain conditions, plasticity is greatly enhanced and that time and the spatio-temporal dimensions of visual experience powerfully determine the level of adaptation both at the functional and the structural level. We further show that other brain regions, cortical and subcortical, participate in the process. In the other we develop a comprehensive theory of image understanding and adaptation, and implement it in a computer model. The theory's central tenet is that image understanding in organisms proceeds directly from adaptively seeking whole images and not via a preliminary analysis of elementary features, followed by object reconstruction. The image seeking adaptive network ;ISAN; computer implementation of the model demonstrates that the theory is viable and has potential for practical applications.**

1. EMPIRICAL OBSERVATIONS

1.1 APPROACH OUTLINE

The general principle that has guided our empirical research has always been that ultimately any neural system has to serve behavior and that behavior serves survival. Evolutionary selection makes it so, even though, once systems have developed, their use can vary considerably from the intended one.

Our research on natural vision has thus taken into account that image understanding requires, from the start, the ability to seek space varying images of interest, integration with the sensory-motor system and plasticity which is necessary to adapt to specific environments. As a consequence even though the specific experiments discussed here address adaptation in neural networks dedicated to vision they are designed in the larger framework of our previous and future work which targets vision as a system. In the same fashion the theory and model we introduce here are meant to be general, and not just ad hoc constructs to explain our results on plasticity.

The experiments and results described here provide, in our opinion, an important missing link that allows integration of many facets of vision research that have remained unconnected. This integration is possible because our methodology has enabled

us to design plasticity triggering experiences {PTEs} which powerfully engage the adaptive capabilities of neural networks. Moreover the characteristics of the experiences are such that unique functional properties are produced in nerve cells, making the identification of traces of experience unequivocal. The nature of the encoding process and the structure of knowledge representation are thus revealed.

1.2 A method to search for memory traces

To study how visual knowledge is represented in the brain it is not sufficient to locate where memories are-- one must also study the functional properties of neurons that encode them. One must further be able to recognize which functional properties are the result of experience and which ones pre-existed. To this end an experience is required that is simple, so that its traces can be easily identified; natural, so that there are no questions of abnormal influences on neural tissue; and finally the experience should not occur in nature, so that its neural traces are uniquely distinctive. These contradictory requirements are surprisingly easily accommodated. We use a behavioral task in which an image signaling danger is shown to one eye only. The animal can avoid the danger {a mild shock to a forearm} by flexing a leg. Leg flexion turns off the danger symbol and turns on an image that signals safety to the other eye. As an analogy

imagine wearing glasses with a red filter for one eye and a green filter for the other ;used to see movies in three dimensions; and arriving at a red light. That red will be seen only by the eye with the red filter. After stopping for a while, ;the correct response; the light will turn green and one can go on, except that the green light was seen by the eye with the green filter! These experiences are simple, completely normal, but they never occur normally unless one wears special glasses. Our animals wear special glasses only for a few minutes each day during training and live in a normal environment for the rest of the time. Those few minutes have disproportionately powerful plasticity triggering effects. We call these experiences PTEs.

1.3 Behavioral training.

Beginning with the fourth-seventh week from birth, the exact time being determined by the size of the kittens, animals were trained in one of two tasks. In one task, as briefly described above, the kitten was presented with a visual pattern consisting of three horizontal bars delivered to one eye only. This stimulus signaled that a mild shock would be delivered to one of the forearms unless the correct response, consisting of that leg flexion took place within half a second from stimulus onset. Forearm flexion, i.e. the correct response, was followed by the disappearance of the danger stimulus and by presentation of a safe stimulus ;three

vertical bars; to the other eye. The animals received about eight minutes of training per day. The animals spent the remainder of the time, about 23 hrs and 50 minutes in large, multilevel cages with their littermates where they engaged in play and exploratory activities. Much attention was payed to them by the animal caretakers, by the experimenter, students involved in the project and the veterinary persnnel. These remarks are not just meant to indicate that we took good care of our animals, we did, they emphasize that the training session involved but a small fraction of the animal time and activities. The complexity of the visual images and of the motor sequences encountered in these activities vastly exceeds that encountered during training. It is thus quite remarkable that the adaptive effects induced by the training procedure should be so large as to overshadow all else. The other task, also unique, consisted of simply alternating two patterns, vertical bars for one eye and horizontal bars for the other, at different periods. For example in one shedule one eye would see a pattern consisting of two or three black vertical bars on a white background for 400 milliseconds, S1, while the other eye was in the dark. Then S1 would go off and the other eye would see two or three black horizontal bars, S2, for 400 milliseconds. This shedule would be continued for about eight minutes a day also. The rationale for the odd figure of eight minutes a day is that on one hand this time proved to be sufficient to generate clear cut phenomena of neural adaptation and on the other hand it also

proved to be the limit of our animals patience.

1.3.1 Electrophysiological results.

We recorded the activity of single cells in the following cortical and subcortical areas: visual cortex area 17, Claire-Bishop (a cortical association area that receives however a direct input from the lateral geniculate nucleus), sensory-motor cortex, hypothalamus. Purpose of the recording was to study **response characteristics** of nerve cells with particular reference to differential responses to bars of different orientation moving in a direction orthogonal to the length of the bar. We will refer to **orientation sensitivity** (see refs. 10, 13 and 14 for more detail) as the orientation of the bar that engenders the largest response when the bar is moving in the preferred direction. The **preferred direction**, in turn, is that direction of motion, for any object, that activates a given cell most. Cells were also tested for **binocularity**, that is the strength with which each eye by itself could activate a same neurone. This measurement yields values from 1 (left eye only) to 7 (right eye only). Perfect binocularity is denoted by 4. Single neurones were also tested for **polysensory responsivity** and that is the ability to respond to non-visual stimuli, such as auditory and tactile.

The picture that emerges from these recordings is that the

functional properties of nerve cell are isomorphically shaped from the experience. What this means is that the dimensions of the experience, be they concerned with shape, orientation, ocularity or other sensory systems, result in analogous functional properties of nerve cells. The implication is that a simple parallel storage mechanism is at work and that a sophisticated retrieval mechanism, possibly associative in nature is then responsible for memory partitioning and retrieval. Cortical and sub-cortical "real estate" are also allocated proportionally with the importance of the experience. Fig 1 illustrates these concepts.

1.3.3 Behavioral consequences.

Our maps of the sensory-motor cortical areas revealed that, after our animals received training of the first type of plasticity triggering experience we described in section 2.3, the representation of the trained forearm was several times larger than the untrained one. Since Penfield studies of the cortical representation of the body sensory surface and of the motor map it has been known that these two maps are topologically, but not topographically, accurate. Body parts occupy cortical real estate in direct proportion to their sensory and/or motor sophistication (10, 14). We thus expected that our animals would exhibit considerably more skill and a preference in the use of the

cortically over-represented foreleg as compared to the normal one. Normally cats exhibit no paw preference, in fact only human beings have clearcut hemispheric asymmetry and handedness. The quantification of skill proved extremely difficult and, even though it was obvious to the observer that the cortically over-represented forearm was considerably more skillful, we abandoned it. On the other hand the quantification of preference of use of one foreleg vs the other is relatively simple and can be reliably performed. The only problem with this type of measurement, we quickly discovered, is that it is task sensitive. Consider the case of a right handed human-- most task are performed cooperatively by the two hands-- only some tasks preferentially invoke the right hand particularly when tools (pen, screw driver, etc.) have to be used, but even here the left hand helps. This preamble is to explain Fig. 2. In this figure we graph forearm preference in percentage points during the performance of a number of tasks. The tasks vary from simple visually or tactually elicited placing of the forelegs on a support surface, to reaching out to capture and manipulate a play object. Visually guided placing is separately shown in Fig. 2. because it deserve special mention. The performance of this test is done by simply holding the animal upside down with the eyes opened and the forelegs free near, but not quite within reach of, a table. Almost immediately the animal is moved closer to the surface and, when the animal sees that it can be reached, it

will reach for it to support itself. Normally a cat will reach with both forelegs simultaneously or one leg will go first, but there is no systematic preference. In our animals there was a definite, statistically significant preference to use the foreleg that had been trained in task number one. The result is quite strong and surprising because the nature of the task is such that even though it does involve visuo-motor coordination it could be considered to be one of the innate visually guided postural reflexes. Clearly it is not. The next figure shows differential preference of use for various tasks for the trained vs the untrained forearm. To parcel out the effect of age at start of training the population is divided in three groups: young, medium and old. We were very surprised to find that there were tasks for which use of the untrained forearm was actually preferred. Our expectation had been, naively, that if hand preference was produced then that hand should be preferred all the time. Closer inspection of this set of data, also statistically significant, however does reveal a pattern and that is that the trained forearm was preferred whenever the task involved flexion motion. Thus it would seem that the enlarged motor representation of the trained forearm from which flexion could be elicited by electrical stimulation does set up a preference for tasks that have an intrinsic affinity with its own capabilities. We could call this a learning predisposition toward a certain class of tasks. We consider these results to be

important in that they suggest ways in which learning predisposition to certain classes of tasks could be engendered in humans. From a more theoretical point of view this the first time, to our knowledge, that hemispheric asymmetry and handedness have been induced so pervasively by such a simple procedure.

1.4 Relating neurophysiology to learning theory

There is no doubt that neural plasticity is much greater during development than at any other later time. However the state of development per se is only one element-- the other equally and possibly even more important is the nature of the experience that triggers the plastic response. It is the nature of the experience that determines how extensive neural restructuring will be. The learning task we used is the result of many refinements and is quite powerful as a plasticity triggering event. It was not designed using learning theory concepts, however it can be interpreted using learning theory ideas. Learning theory also accounts for its effectiveness. Quite simply learning theory, e.g. Rescorla-Wagner's, states that association between two events $\{E1 \text{ and } E2\}$ is promoted when $E1$ predicts $E2$, however $E2$ has to be "surprising", that is informative. In fact the growth in strength of the association is proportional to this surprise factor. Thus the theory concentrates on $E2$. In our experiments these factors make the training unusually

powerful--- a) it fall into the class of **sensory preconditioning** because E2 is initially unpredicted by E1. b) presenting the stimuli with goggles minimizes **contextual overshadowing** thus further increasing associative strenght. c) the task also has the properties that produce **superconditionig** because, in animals with frontally located eyes e.g. cats and human beings, if one eye sees E1 then the prdiction is that the other eye should see E1 also!! This makes E2 **more surprising**, in fact incredibly so. In this light it is easy to see why the brain changes produced by this PTE task are so massive, precise and long lasting. **These concepts have implications for the training of humans when it is especially important to minimize training time and/or obtain particularly powerful and long lasting results.**

1.5 Summary

Much remains to be discovered before we fully understand the adaptive networks that perform image understanding in organismic brains. However what we have discovered regarding adaptation, combined with concepts developed from some of our previous work and from work of others is sufficient to generate a preliminary theory of image understanding.

The theory is radically new and we have a computer implementation that demonstrates its feasibility.

We believe that practical image seeking devices could be built

using the principles outlined in what follows.

2. A NEW THEORY OF VISION

2.1 THEORY OUTLINE

Even though much progress has been made concerning image understanding in animal and machine vision researchers are still quite far from having formulated a theory that is susceptible of computer simulation or even one which is generally accepted. In fact researchers are still discussing what it is that the visual system actually does! (1,3,4,6).

At present image understanding appears to be a problem of inordinate complexity due to the extreme variations in size, orientation and spectral composition of images generated by any one object on a retina or in a camera. After detection of the elementary features in the image, recombination into **meaningful** objects and separation from background is necessary-- here the possible number of recombinations explodes. In some ways the problem is similar to that of an analytical chemist-- the breaking down of molecules into atoms requires **subsequent resynthesis**. Naturally as the complexity of the molecules to be investigated increases so would the difficulty of the analysis, until the combinatorics would make it impossible.

However a chemist does not indiscriminately break down to atoms an unknown large molecule-- rather the molecule is broken only to the largest known pieces! Detection of known parts of molecules, themselves molecules, can be absurdly simple using specific reagents.

We aim to demonstrate that this is precisely how the visual system operates and that is by seeking images of interest directly and by breaking down unknown images to known ones. Thus only rarely, and mostly during development, would the visual system resort to elementary feature analysis.

2.2 PROLEGOMENA

Although there remains much room for discussion, a reasonable working hypothesis, to us at least, concerning the basic requirements of an animal, e.g. a cat or a bee, vision system would be as follows:

- 1) to enable motion in an environment without undesired collisions
- 2) to seek, or run away from, objects of interest
- 3) to develop, via learning, an internal representation of the environment which is adequate and sufficient to enable the animal to travel from home base to targets and return.

Extremely minute amounts of neural tissue, such as the brain of a

bee, realize navigational systems that are astounding in sophistication and precision.

2.3 Some heuristics in theory formulation

In formulating a comprehensive theory of vision we are guided by four elements-- one has to do with empirical data, our own and of others, generated in the field of neuroscience and behavior. The second element is that of the three functions postulated above the second one is either the most important, the most self sufficient and/or the first to appear on the evolutionary scene. The third guiding principle, really, a tenet of modern evolutionary thinking, is that for a function to develop as we now know it, e.g. flight, selective advantage must be present at each and every step, large or small from the very beginning.

From the point of view of vision this means that even primitive vision, had to be useful from the start. Given that objects of interest are first far away then very near, with consequent major changes in image size, size constancy and translation constancy have to be properties of the basic architecture, or the vision system would have been unable to evolve from simple intensity seeking, which appears first, to image seeking.

To put the argument a little more strictly from the selective

advantage point of view, and postulating random variation, the appearance of light sensitivity in one sensor in some animals would provide them with a selective advantage over the ones that don't have it, so that the latter group would be selected out (a heat seeking missile is better than one that doesn't). The random appearance of more light or heat sensors would provide more selective advantage because the function would survive damage to one or more sensors. Transition to patterned light tropisms would then be possible as long as the next variation provides a selective advantage.

Thus some object recognition, no matter how simple the object, has to be possible regardless, within limits of course, of distance, orientation or spectral composition of the light (the spectral composition of light outdoors varies depending on the time of day or simply sky direction even on a clear day). Only then these animals would win in the struggle for existence over the others.

Subsequent random variation would continue this process in the same fashion with greater advantages accruing to those variations that favor survival in the specific habitat or domain of action of a given specie.

I have spent some paragraphs on this issue because our knowledge

of vision systems is incomplete and theory development critically depends on heuristic principles ;see for example J.Feldman considerations of time factors in animal vision and the impact they should have in the design of machine vision (23); and working hypothesis which need to be made as explicit as possible. The evolutionary argument detailed above is an extremely powerful heuristic in neural circuit design because it says that a new component is simply out of the question if it requires another component to make it useful which will come later. Later, in evolution, means millions of years and a useless variation would not be selected for. In this light complex circuits are simply out, the probabilities are just too much in favor of short and simple circuitry. Lastly a theory should be susceptible of computer simulation, partly because of the important practical applications that such an implementation would have, but also as a necessary test for logical consistency. In our experience quite often ideas that appear eminently reasonable to oneself and others turn out to be unworkable.

3. INITIAL STEPS TOWARD A COMPREHENSIVE NEW THEORY OF VISION

In attempting to build a new theory of vision it seems appropriate to remember, with some humility, that vision reserch, in animals and machines, is and has been the arena of giants. Further it's in the nature of experimental science to proceed,

generally, from the most observable to the least observable and that, quite often the most observable is also the most fundamental.

Thus a new theory of vision must perforce rest, not just on one's own work, but also on the work of others to such an extent that it might be difficult to actually see what is new and different. However, apparently minor differences can be all important, whitness the impact that minute discrepancies in computed vs. measured atomic masses had on nuclear physics, beginning with nuclear reactors, the atom bomb, etc. Thus while conceptually this new theory is radically different from previous ones, it should be noted that structurally and functionally we make appeal to data and relationships which are not on the surface fundamentally different from those described by other workers. Furthermore, and most importantly, some of the work we will make appeal to is very old and somewhat forgotten, but in our opinion, falls in that class of most observable therefore most fundamental phenomena we referred to above.

In an initial beginnig toward a comprehensive theory of image understanding (10,19) we noted that:

- 1) the vertebrate retina has remained unchanged, except for some ecologically significant adaptations, since it appeared on the evolutionary scene about 400 million years ago and

interpreted this fact as an indication that, structurally and functionally, no variations which must have taken place as they did in invertebrates, proved to be advantageous. Thus we are in the presence of a structure which must be nearly optimal for image understanding regardless of the type of brain possessed by a particular species. Vertebrate brains, and most other structures, have undergone considerable evolution and exhibit substantive differences from one species to another. More particularly the vertebrate retina serves equally well animal species that have and those that do not have a geniculo-cortical system.

2) at each processing layer lateral inhibitory phenomena can be clearly demonstrated. There are two inhibitory networks in the vertebrate retina-- one served by the horizontal cells in the outer plexiform layer and the other by the amacrine cells in the inner plexiform layer. These layers are located between the photosensors, rods and cones, and the bipolar cells and between the bipolar cells and the ganglion cells which originate axons that connect the retina to the brain. Animals with a geniculo-cortical system have another layer of lateral inhibition in the lateral geniculate nucleus.

3) it is now generally recognized and accepted that numerous efferent fibers, possibly many more than from the eye itself, reach the geniculate nucleus from cortical areas--

controlling and modulating lateral inhibition. We and others, have data showing that efferent control extends to the retina (7,18). Even though not all investigators believe that all vertebrates possess efferent control to the retina it is unquestioned in many species. In any case, for our theory, it is sufficient that some form of efferent control be present in one of the lateral inhibitory networks of the input pathway e.g. the lateral geniculate nucleus, which is unquestioned or the retina, for reasons which will become apparent later.

From these stage setting considerations, based on much data of ours and of others (19,20), we conclude that lateral inhibition must be at the core of retino-geniculate information processing, an idea which has influenced vision research practically from its inception. However, based on considerations 1 and 3, and again guided by much empirical data (17,18), we rejected the generally espoused interpretation that the function of lateral inhibition is simply and only that of enhancing contrast, that is to amplify more high spatial frequencies in the image.

The incontrovertibly demonstrated efferent pathways combined with attentionally driven changes in receptive fields shapes at the lateral geniculate level led us to ask the following question-- is it possible for the higher centers, i.e. visual cortex and

other structures visual and non visual, to modulate meaningfully ;we know that modulation takes place; lateral inhibitory functions so that each layer, instead of indiscriminately amplifying contrast, would selectively amplify an image as a whole? We answered the question affirmatively by demonstrating, with the aid of a computer simulation, that inhibitory functions exist such that certain images in the visual array are amplified while others are not, even when the elementary features of the different images are the same, ;vertical, horizontal and diagonal edges; as shown in a previous report (19). In our opinion this finding is quite remarkable because it shows that a vertical edge is enhanced or not depending not just on its physical properties, but on its **meaning** which is determined by its belonging or not to an object of interest. That in turn is determined by other systems ;not by geometrical properties except in the case we are interested in geometry; that have phylogenetic and ontogenetic knowledge, via efferent control pathways to lower input structure. Referring back to our mythical chemist hydrogen and oxygen, when combined, behave differently-- a test for a property of water will not bring out the hydrogen that is present in paraffin.

In this way the visual array is not broken down to elementary features from which objects of interest have to be reconstituted -- objects of interest are selectively amplified or

enhanced directly from as early as possible in the processing chain and that chain is very short-- reaction times are critical in the struggle for existence, and reaction times are astonishingly short, in animals, if one thinks about how slow neurons really are ;it can be argued that a nerve cell needs at least 10 milliseconds to reliably determine the firing rate of its input.

The previous report (19) left many questions unanswered or only partly answered. For example-- are these lateral inhibitory functions physiologically meaningful? ;a negative answer would not diminish the potential usefulness of the algorithm for machine vision;-- how do these functions come about in animal brains, that is are they genetically built in, acquired during development or are they learned and if so how?-- whether built in, acquired, or learned, are these functions optimal or could we, mathematically, develop better ones?-- what are the limits of the algorithm in terms of complexity of an image?-- given our initial argument that at least some size, orientation and spectral constancy should be present *ab initio*, is it, and if so how much?-- In this section of the report we will review previous theory development, extend our theory significantly, and thereby attempt to answer some of the questions posed above

4. IMAGE SEEKING ADAPTIVE NETWORKS-- ISAN

ISAN's core structure is derived firstly from what seems to be universally present in receptor surfaces either directly or at the next synaptic layer and that is lateral inhibition. The second core feature is the efferent system that modulates it. Lateral inhibition has been studied in the vertebrate retina, in the invertebrate eye of Limulus and for the auditory and skin receptor system by many workers (20). Some of the most precise mathematical formulations of these lateral interactions (even though we refer to lateral interactions as lateral inhibition, partly for historical reasons, lateral interactions is in fact a more appropriate name because lateral excitation is often present) have been summarized and compared by Ratliff (20) who concluded that essentially the formulas he discussed, six in all, were comparable in their action.

Neither Ratliff or any of the investigators whose work he presents, consider efferent control and its action on retinal function. Interestingly even such a simple, primitive and evolutionarily ancient invertebrate as Limulus has an efferent system to the eye one of whose functions, most certainly not the only one, has been elucidated recently. There are authors who have contributed to the considerable literature on efferent control systems in general and to the retina and lateral geniculate nucleus in particular (7), but they have not

generated, to our knowledge, a global functional theory regarding its function. Theories concerning possible functions, of course, abound-- no theory we know of however provides the mechanism that would implement them.

Most of these theories perforce postulate functions related to attention.

ISAN's theory, supported by a computer model, parallels these ideas.

There are some important, in fact critical differences however. One of the principal ideas in ISAN, at real odds with most other approaches, is that here is no abstraction process. A single neuron firing as an indication that there is an edge, or a texture region in some region of space is an abstraction of that information because the imperfections, or better yet the details, of the edge and/or region would be lost.

That is not the way we see.

Instead we subjectively experience that all of the information contained in the visual array is available at all times except that some parts of it, objects of interest, are more salient. Our hypothesis is that the subjective impression of effortless

seeing and locating in space ;in general of course; objects first and then details and or background later is the direct outcome of equally simple hardware that directly sees and locates objects of interest by selectively amplifying them more.

Quite simply more gain means that higher or lower firing rates are obtained by the active elements than would be accounted for by the physical properties of the image. This is an extremely important effect and it accounts, as we will show, for the peception of clear edges in situations where edges are obscured by noise or even absent.

Only if the details of the object become of interest are they analyzed further and effortlessly because, as part of the object, they have been amplified more than possibly identical details which are part of un-interesting objects.

LSAN provides a simple mechanism, lateral inhibition, that can perform selective amplification of images of interest, i.e. the gain factor for these images is larger than for those that are of no interest to the system.

It must be pointed out that images can be thought of, from a Fourier analysis standpoint, as if they were made up of sinusoidal spatial frequencies and that, from its discovery,

lateral inhibition has been thought of as a mechanism for contrast enhancement, i.e. capable of amplifying high spatial frequencies to a greater extent than the lower ones.

However the result of making the assumption that the details of the inhibitory function are to be ignored and that, ideally, its shape can be equated to a Gaussian results in a processing surface that indiscriminately enhances contrast whether it belongs to objects of interest or not.

Inordinate requirements are then put on higher structures to reconstruct and disambiguate objects from each other and from the background.

It is our contention that the details and variability of the inhibitory function, also remarked upon from the very beginning by most workers (5), must not be ignored as it is the whole function, details included, that determine what spatial frequencies in combination with others are selectively enhanced.

Different gain for different whole images can in fact be obtained by appropriate efferent modulation of the inhibitory function.

4.1 THE ARCHITECTURE OF ISAN

The basic structural architecture of ISAN is shown in Fig. 3.

This is the architecture of the **theory**, not of the retina or the lateral geniculate-nucleus. There are some differences in these structures, between species and even between individuals of the same specie because of phylogenetic and onthogenetic adaptation (10).

As previously remarked (19) not all cell types are considered in the theory. These parallel channels, unquestionably provide additional capabilities which we will attempt to elucidate at some future time. It is our conviction however that ISAN's three layers capture the fundamental function of the vertebrate retina and geniculate nucleus.

This would seem to be an improbable statement an the face of the fact that it can be readily seen that there are no horizontal or amacrine cells in the first two layer, the retina, nor inhibitory interneurones in the third ;the geniculate¿. However these cells are necessary, in vertebrates, to implement a **better** lateral inhibitory function than would be possible say in the eye of Limulus, an invertebrate in which lateral inhibition is implemented much as in our model, and that is without the aid of interneurones. Interneurones expand the range of lateral inhibition and also, given that an interneuron generates

synapses which can be only excitatory or inhibitory, are necessary to enable lateral inhibition and excitation.

We could, of course, have simulated all of these cells but it seemed more expedient, due to the fact that silicon does not have the limitation of flesh, to endow the neurons with perfect linearity and to allow synapses to be excitatory or inhibitory as required even though they originated from the same elements. These simplifications do not affect the outcome and have been used by other workers e.g. Ratliff in his simulations of the eye of Limulus.

Let us now look at ISAN's structure in detail.

First we notice that, except for the labels, the three layers are identical. Second, each cell ;not all connections are fully drawn; can influence the firing of any other cell in an inhibitory or excitatory way. This is denoted by a small circle, which represent the synaptic bouton, which is half empty and half filled to denote excitation and inhibition respectively. Thirdly each synaptic bouton in turn receives a pre-synaptic contact from efferent axons, that is from axons that originate in the brain.

When pre-synaptic inhibition was discovered there was some initial skepticism because it seemed that too much precision was

required of those mechanisms that are responsible for wiring the brain. These connections have however been verified with the electron microscope and there is general acceptance of their existence. There is also agreement that pre-synaptic inhibition, or excitation, is a powerful modulator of synaptic efficacy in that transmitter release is under its control.

In this fashion lateral connections have a preset or wired in static strength, a default mode of action, represented by the synaptic bouton's weight. Activity on the pre-synaptic terminal can however dynamically change both the sign and the strength of synaptic action thus enforcing, albeit temporarily, a different goal.

It must also be pointed out that the lateral interaction function, whether built in or imposed from above, applies to each and all cells in the network, (excepted at the edge of the networks where, unavoidably, there is no lateral inhibition from the outside), thus the same action takes place at all points. Thus translation constancy is assured. More will be said later about size and rotation constancies.

The basic philosophy of this model is that local changes in the nature of lateral interactions are produced by efferent pre-synaptic inhibition to implement global goals, thus the

control function is broadcast to each and every cell from above.

The idea that global actions can be exercised on large portions of the central nervous system is not new-- e.g. the reticular formation globally activates or inactivates cortical areas. However these effects are relatively non-specific and quite different from the type of action envisaged here. We will show further on that total precision is not required and that a substantive percentage of mis-connections can be tolerated.

Again, rather than proliferating connections to separately account for pre-synaptic inhibition and excitation, we allow these terminals to cross the zero boundary so that, when required pre-synaptic inhibition can become excitation or vice versa.

4.2 MECHANISM OF ACTION FOR SELECTIVE IMAGE AMPLIFICATION.

Quite simply if we agree with the generally accepted idea that lateral inhibition de-amplifies the spatial frequencies defined by its Fourier transform, as cogently argued by Ratliff (20), then lateral excitation must amplify the spatial frequencies defined by its Fourier transform.

More generally, by properly selecting a lateral interaction function that contains excitation and inhibition as defined by

the Fourier transform of an object we can make a processing layer, based on lateral inhibition, amplify that object more than others, much as in a hologram.

However the retina does not perform a Fourier Transform.

What the retina does is an instantaneous sum, moment by moment, of all the lateral interactions that are taking place locally and in parallel.

Just as Hartline&Ratliff and others did, we referred to the Fourier transform simply as an explanatory device. There is a wealth of mathematical transformations and others could have been used. Our preference for the Fourier transform is that it is closer to the physical world and it is therefore easier to visualize its action.

The sum of all the lateral interactions moment by moment can be quantified by the following formula:

$$x(i) = E(i) - \sum_{j=1 \text{ to } n} k(i, j)x(j)$$

This formula describes a system of simultaneous linear equations, where each $x(i)$ represents the output of each element, each $E(i)$

the excitation impinging on it and the $k(ij)$ the strenght of each lateral connection. The results of using this approach are physiologically more accurate and, mathematically quite interesting.

Using abbreviated matrix notation

$$Ax=b$$

it is easier to notice that there are three components to the expression-- the b vector represents the inputs or $E(i)$ in the previous equation, the x vector the $x(i)$ or outputs and the A matrix the inhibitory coefficients. We further note that if two of these elements are known the third can be computed-- thus:

- a) Given the inputs and the outputs the coefficients can be determined;
- b) given the inputs and the coefficients the outputs can be computed;
- c) finally given the outputs and the coefficients the inputs can be reconstructed.

We have belabored these aspects pertaining to systems of linear equations to prove a point and to ask a question. The point is that the transformation code is complete because the image can always be reconstructed from the output $\{c\}$. The question, suggested by $\{a\}$ is as follow-- given, as input, a visual array that contains two or more objects can we, arbitrarily, set up an

output in which one of the objects is amplified more or less than the others, solve the equations and find the right coefficients as in ;a; so that in the future we could, trivially, do ;b;?

We realize immediately that not all systems of linear equations have solutions thus there are strong limits to the word arbitrary used above, further not all existing solutions might be biologically meaningful. Referring back to our chemical analogy-- there is not, and interestingly it's not necessary or even desirable to have, a specific reactant to each and every possible chemical.

Lets make the above very explicit by actually looking at the equations in expanded form for a five cells, one dimensional, network

$$\begin{aligned}
 x(1) + k(1)x(2) + k(2)x(3) + k(3)x(4) + k(4)x(5) &= E(1) \\
 k(1)x(1) + x(2) + k(1)x(3) + k(2)x(4) + k(3)x(5) &= E(2) \\
 k(2)x(1) + k(1)x(2) + x(3) + k(1)x(4) + k(2)x(5) &= E(3) \\
 k(3)x(1) + k(2)x(2) + k(1)x(3) + x(4) + k(1)x(5) &= E(4) \\
 k(4)x(1) + k(3)x(2) + k(2)x(3) + k(1)x(4) + x(5) &= E(5)
 \end{aligned}$$

This elementary system of linear equations, easily expanded to

two dimensions, precisely describes the computation performed in each layer of ISAN. As such it is not very different from what has been accepted by many Authors. It differs fundamentally however in how it is used by not ignoring efferent control modulation of the k and by realizing that the lateral interactions can be locally adjusted to achieve global goals i.e. image seeking instead of indiscriminate contrast enhancement. It also makes again obvious how utterly simple the computation performed by the brain is and how fast-- given that it is done in parallel.

4.3 ORIGIN OF EFFERENT CONTROL FUNCTIONS.

In a previous report (19) and at the beginning of this one, we argued that, because of time constraints, the vision system could ill afford an atomic analysis of the visual array. We claimed that it would be enormously faster and easier to directly seek objects of interest. We then set out to show that a lateral inhibitory network in which the lateral inhibitory function could be retuned by efferent control from some initial values, could be capable of amplifying certain objects more than others.

We further argued that efferent control is, biologically speaking, a necessity because it guarantees that information, visual or otherwise, is processed in the context of its meaning

as early on as possible and that it is primarily up to the higher centers to know about meaning and to inform the lower centers of what needs to be enhanced if it's out there.

There is no question in our mind that such a system is superior, in terms of reaction time, to any other one. Just as our chemist does, precious time is spent not on the final reaction, but in the previous preparation of the specific reactants.

Does it then mean that the visual system has to solve linear equations to develop proper control functions? We suggested that some of these functions, probably the most important ones for animals such as frogs, are built in by the genome, either in the higher centers to be sent down on specific occasions e.g. looking for a mate, or permanently embedded in the retina i.e. looking for flies. Other functions would be acquired during development and later on with learning.

Here we extend our theoretical framework in three directions. In one we explore, from a mathematical standpoint, the limits of systems of linear equations as a method to uncover control functions for selective amplification of objects of arbitrary complexity. In the other we develop a working hypothesis as to how organisms develop control functions and show that it works in the ISAN network. Finally we define and construct a cortical model

that accounts for the plastic phenomena we have observed at the cortical level.

4.3.1 A mathematical approach.

As delineated above a system of linear equations can be described, in simplified matrix notation by the following formula

$$Ax=b$$

where, in our case, b represents the vector of known quantities ;input¿ and x the vector of the unknown ones ;output¿. We were surprised to discover that there are very stringent limits, given a certain set of inputs, with regard to the set of outputs that has a solution in terms of inhibitory coefficients.

Given an input containig a number of objects that have identical contrast ratios, with e.g. a square among them. We construct an output in which the square has larger numerical signal values than the other objects. We then solve the equations, find the set of coefficients and do the mirror process, that is plug the image and the coefficients in ISAN to see if indeed selective amplification for the square is obtained.

Needless to say the system works, if there was a solution for

that particular object.

Often however a solution cannot be found and we have been investigating a mathematical method that, in these cases, will still find a least square approximation to the desired solution. These studies are at an early stage, but we are convinced, and some experiments with the method described below lend support to this conviction, that it is possible to analytically compute control functions to selectively amplify objects of considerable more complexity than the ones we have been using.

4.3.2 A possible biological approach (1).

A mathematical solution to these problems would be highly satisfying to the intellect and would have possible applications in the design of novel, parallel architectures for computers dedicated to image understanding. Much as in the case of flying machines once the basic principles are understood it is usually possible to greatly outdo those mechanism which have, after all evolved through random variation.

Even at this early stage of development ISAN can outdo a human observer in "seeing" an image heavily embedded in noise!

It seems however unlikely, at least to us, that the basic

circuitry of the visual system can perform highly complex and precise mathematical transformations. It seems enormously more plausible that relatively simple circuitry performs noise resistant operations such as summing and subtracting, or implements functions such as negative feedback which are considerably insensitive to variability in the components.

These types of operations have in fact been observed and described in neural structures.

How are new control functions built?

There many possible scenarios for action with regard to vision. Three are of particular interest in the context of this exposition.

In one a specific, known object is being sought. In this situation, the ideal case if there is any merit to ISAN's theory, the lateral geniculate nucleus and possibly the retina are preset to selectively amplify that image. If it is present in the visual array it will create an area of high activity in the processing layer which will induce, quite simply, orienting reactions of eyes, head etc. as necessary to bring this area in the center of the visual array for further inspection if needed. Appropriate behaviors will be then be elicited as the position in space of

the object is, by knowing the position of body parts, known. If the object is not present in the visual array other systems would be invoked.

In the second case a known object, which is not being sought, enters the visual array. There are two possibilities now-- the visual system might be actively and strongly seeking some other object in which case the new image might go completely undetected ;just as we would never know that a previously off TV station has returned to the air if the set is tuned to a different channel; or, either because the tuning is low or the signal strong, the activity generated is high enough to reach the threshold that induces an orienting reaction. Somehow, regardless of how previous experience is represented in the brain, the proper memory traces have to be activated. We know from our own experimental results on early visual experience (10,11,13,14), that receptive fields shapes of nerve cells in visual cortex closely embody relevant aspects of the experience. These cells will respond more strongly than others and will inhibit other cells with different response characteristics and, most importantly, by activating the efferent pathways tune the input path to their own selectivity making it more selective for the same object ;the reader familiar with electronics and phase locked loops might think of the signal capture capability exhibited by these devices as a helpful analogy. The nature of

this feedback is positive and the object would quickly stand out for further action as necessary.

The third case is that of an unknown object. Here also there are two possibilities-- the object is unknown, but all of the parts are known, or the object is unknown and some or all of the parts are also unknown. We encounter at this point an epistemological problem that requires a definition and a brief digression. The digression is really a reminder that animal vision goes through a critical period of development characterized by tremendous plasticity. During this period the visual system literally wires itself together and the properties of receptive fields of neurones in visual cortex are defined and shaped ;or re-defined and re-shaped; by experience in rather direct ways. Clearly evolution has not discovered the best set of elementary features for general image understanding, instead a domain specific set is ontogenetically acquired by each individual!.

The definition is really a problem-- what is a part in an object's image? What set of parts would be best in a certain environment ;domain;? How are parts represented and accessed in memory? How are the relationships between parts represented? These are very complex problems for which we think the brain has found a simple and elegant solution.

A part, or a component, is a fraction of the whole which usually has been detached along some natural lines of cleavage ;the natural lines of cleavage of the brain after fixation were used as a method for classification by the very early anatomists-- thus the name cortex, which means bark, was given because it detaches as bark does from the rest of the mass.

This definition however only sidesteps the problem and does not solve it. What is natural in geometry might not be natural at all from the point of view of the relative frequencies with which groups of elementary features (parts) occur in images which are common in a certain environment. Natural lines of cleavage might be determined by function or by meaning. In fact when we look at images we can partition them in very many ways whose number depends only on our imagination. Needless to say parts are in turn made of parts.

A possible solution comes to mind by referring to the reminder above, and that is the almost iconic shaping of visual receptive fields that takes place during development, by recalling our hypothetical chemist and by making connection with a fascinating branch of computer science which deals with the statistical properties of text at increasing levels of joint probabilities (Bennett). This field is often referred to as random language generation. What is particularly interesting about it is that

fourth order random text sounds like, stylistically, the original text from which the statistics were derived. Storage space requirements for the joint probability arrays, unfortunately, increases exponentially with higher dimensions.

Remarkably Hayes has pointed out that there is really no need to compute and store all the joint probability tables, impossible in any case for the higher orders, because they can be trivially reconstructed from the original text in which they are contained to begin with in their most compact form!.

In this fashion, it can be hypothesized, does the visual brain represent knowledge so that it can be most readily utilized.

Visual information would be stored in an amount sufficient to define the statistical distribution of parts present in the environment. A part can be the whole object or a smaller section that has recurred before with increasing or decreasing order of joint probability, but always trying to use the highest order available. that is the largest part, so as to minimize the reconstruction time.

This long series of preparatory reasonings provide the answer to situation three presented above-- if an object is unknown because all or some of its parts are unknown then that object, or its

unknown part ;no division; is simply stored. Our data do in fact unambiguously show that when a new experience is encountered, even in an adult animal, that cannot be reduced to previous experiences even in its parts then new types of receptive fields are formed.(21)

4.3.3 Origin of seek functions-- biological (2).

We have come full circle to where we started and that is how do seek functions come about in real brains?

The answer, for ISAN at least, is as follow

1) the image of an object is presented to the visual array in its **ground state**, i.e. the lateral inhibitory function in the retina geniculate nucleus is of the **simple gaussian type**.

2) because, by definition, there are no neurons in the cortex that are tuned to this particular object all that could possibly happen ;in any case this is what happens in our simulation; is that the cortical module that sends its output back to the lateral geniculate nucleus reflects back to the lateral geniculate nucleus what it has just received and that is the image as transformed by the basic gaussian function. **The efferent signal modifies the inhibitory coefficients.**

3) the above cycle is allowed to repeat a number of times, each repetition resulting in a more and more selective efferent function.

4) the successive refinement of the seek function reaches equilibrium in about seven to eight cycles at which time the retina, geniculate and the efferent module are locked onto the object and, even though the object has no meaning and is not stored permanently, the proper seek function is available in the cortex to be permanently learned or discarded.

5) notice that a) a completely new object would take about four times longer to be captured than a previously known one b) as long as the loop is unbroken tracking or any other form of interaction with the object would be identical to those available for previously encountered ones. That is there is no way to distinguish freshly created seek functions from retrieved ones in the temporary memory of the efferent module ;much as in a computer; and only by making reference to permanent storage would the system know that the object is new ;namely nothing pops into ones mind as to what it means or what can be done with it;.

Now that we have a simple, biologically plausible mechanism as to the origin of control functions we can address the question of how they are stored in and used by the brain.

5. DEVELOPEMENTAL PLASTICITY AND LEARNING

Critical periods of developemental plasticity of the brain have been demonstrated unarguably (10). During this time massive resource allocation in terms of cortical representations of the body surface (13,14) and of functional and structural properties of nerve cells in visual cortex have been demonstrated (11,13,14,21).

It seems however unlikely, at least to us, that the fundamental learning mechanisms of the developing brain should be different from those of the adult with the exception of degree.

Even if different mechanisms were to be demonstrated to prevail ;we ourselves have suggested that dendritic bundling could be such a mechanism (22); the encoding mechanism must be the same otherwise these memories would become inaccessible later rendering the whole period of visual developemental plasticity useless.

We thus posit that, while quite possibly different ratios of

change might apply to different structures at different ages, ultimately learning involves changing the strenght of functional connectivity between active elements in the brain, thereby changing behavior, either by synaptic modification or by dendritic bundling or by cell growth or cell death. Other mechanisms yet to be identified would not change this requirement whose ultimate effect is to re-route information flow.

When an image is captured, that is the retino-geniculo-cortical pathway has locked onto it, it means that a set of cells in visual cortex belonging to the efferent module that projects to the lateral geniculate nucleus is very active and is enforcing, if not the best possible, at least a very good efferent control function for that object. If the object is new and that pattern of activity were to be remembered, that is stored, it would enable the animal to seek that object with much less delay next time it appears.

The ability to store more than one function would provide further advantages for survival-- therein lies a continuing evolutionary pressure for larger, faster and better memory ;for computers too.

5.1 A model for visual cortex learning-- CORTL

Even though Lashley could not locate the region of the brain where memories are stored there is now ample evidence of experientially induced functional and structural changes in several brain structures. These range from the spinal chord, to the cerebellum, to visual and sensory-motor cortex (10,13,14).

The functional organization of visual cortex, and by this we mean area 17, 18 and 19 of Brodman classification, can be described from several points of view e.g. types and functions of neural elements and their columnar organization.

It has been suggested that these columns are further organized in hyper-columns toward a hierarchically structured system of increasing complexity meant to parallel and explain image understanding at various levels.

For our purpose however, we intend to subdivide visual cortex using a structural and functional criterion often used by neuroanatomists and that is the output terminus of efferent cells.

If one considers output as the guiding element for classification, visual cortex can be shown to contain at least three identifiable modules which are present in different ratios in the three areas. One module generates an output to the lateral

geniculate nucleus, another to the superior colliculus and the last one sends outputs to cortical association areas. This subdivision does not in any way contradict the columnar organization scheme or exclude other efferent pathways i.e. to the the pons (Glickstein).

We have formulated a theory, and implemented a computer simulation of it (CORT), which takes into account known neuroanatomical data and our own results on cortical plasticity. The theory provides a mechanism that explains and formalizes our results, vis a vis current learning theory, neuroanatomy and plasticity effects demonstrated by other workers. It is not yet a general theory of cortical function, that is for the future, in that it only deals specifically with plastic phenomena as posited to take place in the cortical module that sends its efferents to the lateral geniculate nucleus.

Following the approach used a few paragraphs back concerning ISAN lets first state as clearly as possible what the learning mechanism is meant to achieve then how it could achieve it and finally the actual structures that perform the required function. It seems unarguable that learning should modify behavior adaptively for the animal to obtain an advantage in coping with the vagaries of specific environments. Even specific environments i.e. ponds, plains, forests, etc. show variation.

Larger amounts of environmental variation would increase the evolutionary pressure toward greater learning ability in a given specie-- failing that development extinction would become unavoidable. The word adaptively cannot be overemphasized--

How can this be done? It would appear that the minimal learning mechanism, in a specie that possesses sensory surfaces responsive to stimuli that can be two dimensional, e.g. touch or light, is iconic-- clay can **remember** a fingerprint and a simple pinhole at some distance from a bleachable surface can **remember** images. If this is the path followed by evolution however, it did not stop there-- remembering one image, in the fashion of an afterimage, might be useful, but remembering two or more would provide increasing advantages and possibly at an exponential rate. Furthermore, to be useful, even a one image memory system ;it's interesting to note that what seems to be essentially a one image memory system has recently been described in a butterfly; needs to have from the start some perceptual constancies built in as part of the basic architecture. Modern evolutionary thinking emphasizes that each variation has to provide an advantage for a trend to **continue**-- a tenet that has powerful heuristic consequences theoretically.

Based on our own data that shows that functional properties of visual and somato-sensory cortex are directly shaped by the

experience (13,14,19) we theorize, and implemented in a computer simulation, that visual storage is **almost** iconic. Based again on our data which demonstrate that subsequent experiences do not modify previously shaped receptive fields, but will shape new ones if the new experience cannot be reduced to the previous one, we posit that storage space is protected, so that parts of an image which are not new are not stored again simply because they are present in a new image. The new parts however and their relationships to old parts are stored. Our data also demonstrates, and we thus posit, that the **meaning** of an image, from the point of view of actions and outcomes associated with it (10,14), are part of the storage. We have some neurophysiological data demonstrating plastic effects in the hypothalamus, thus we also posit that internal states, are part of the learned modifications ;state dependent learning is a well accepted concept in psychology;. Internal states, actions and consequences associated with a given image constitute its meaning and need to be kept with it, so to speak, as much as the hardware of the brain will allow. The reason for this is that we are convinced that it is the action of the internal states, such as hunger, etc. that will later invoke the control functions that will make the input pathways so exquisitely sensitive to objects, sounds, or smells of interest regardless of the noise that blankets these signals in the real world.

The mechanisms responsible for all these actions are, in the simulation at least, remarkably simple. Output from ISAN feeds into stellate cells which in turn deliver their output to pyramidal cells (after Hubel and Wiesel). Pyramidal cells also receive a direct input from ISAN. The model addresses itself only to the module that sends efferents to the lateral geniculate nucleus. Based on Sperry's data which shows that visual memory is not lost after severing the optic nerve and allowing it to regenerate, we have made afferent synapses to cortex excitatory and fixed in strength. Afferent fibers make contact with stellate and with pyramidal cells. Pyramidal cells also make contact with stellate cells. It is known that some of the stellate cells are excitatory while others are inhibitory. We follow Hubel and Wiesel's scheme and posit that the excitatory stellate cells make contact with pyramidal cells-- inhibitory stellate cells are responsible for lateral inhibition. Plasticity is limited to the stellate cells.

This part of the model is very similar to von der Malsburg's model of visual cortex. There are some important differences however. Firstly maps of receptive fields contain not only excitatory regions, but well defined inhibitory ones as well (H&W). What this means is that the darker parts of an image do not simply represent lack of excitation but are actively detected also. This feature requires that afferent fibers, originating from

off-center retinal ganglion cells make contact with stellate cells that are inhibitory to pyramidal cells. Hubel and Wiesel suggestion (3) that the inhibitory flanks of elongated excitatory cortical receptive fields originate, passively, from the inhibitory surrounds of on-center ganglion cells in the retina and vice versa, must also be operative. We quickly discovered, however, that an active mechanism had to be implemented to achieve sufficient discrimination.

Another difference lies in the way lateral inhibition is implemented. Low level, wide ranging lateral inhibition is evident in our computer maps of cortical cells receptive fields and was also implemented in von der Malsburg model. Its purpose is to prevent adaptation to the same features by nearby cells and, in conjunction with lateral excitation, accounts for the origin of cortical columns. The parameters of lateral inhibition, if it is to ensure a winner take all situation, proved very difficult indeed, to adjust. The reason, quite simply, is that too little lateral inhibition does not produce the desired effect and too much causes the network to oscillate. We believe the circuitry of the brain to be designed in the best possible device independent philosophy and that is, capable to perform its function in spite of large changes in the parameters of the active elements (electrical engineers use this approach also because the parameters of transistors are quite variable from one

another even in the same type.

Various schemes for winner take all circuitry are possible-- in the end we selected to use one that, because of its physiological plausibility, we had used in a previous model (OCCAM). The desired effect is achieved by allowing laterally inhibitory stellate cell to also inhibit the afferent connections-- in this way cells that are less active receive not only more inhibition from the more active ones, but also lose excitation from the input and quickly shut down.

A comment needs to be made about the learning rule used in the model-- it is extremely simple and it assumes that there is a certain inertia in starting synaptic modifications, this makes sense from conditioning experiments and also from some of our results on plasticity which indicates that there are optimum delays to induce plasticity. It is also necessary to give time to the winner take all mechanism to assert itself. At the end of this initial period all cells that already know some parts of the object are active and they are preventing nearby cells from being modified. The unknown part of the object is being recycled through the cortico-geniculate loop working its way toward capture ;the known parts are being recycled too, but because this has happened before the control function is already known. After about seven or eight cycles capture would be achieved for

possibly not, in any case the plastic machinery is now ready and all those synapses that were activated are set to maximum and those that were deactivated are set to minimum. Previously tuned cells remain unchanged-- cells that were active, but untuned become tuned. New control functions either embodying new parts or new relationships or both are thus added to the total.

We have just finished programming the model and have relatively few experiments with it. Given that our requirements were clear it is not surprising however that, in all the instances we tested, the model has performed in such a way as to reproduce those effects of plasticity that are known to occur-- after all it was designed that way. There are however some effects of that design which were not planned for, but happen to be desirable and critical that deserve to be discussed. One of them has to do with how the model handles large images and/or possible damage to its circuitry.

5.2 Large images and/or cortical damage.

All of visual cortex is occupied by the visual array at all times and even typical objects under relatively close inspection e.g. a word processor screen can subtend 30 to 50 degrees of visual angle. Remembering that the visual field is considerably magnified at the foveal projection the result is that such an

object is spread out all over visual cortex. Relatively small movements of the head can generate large percentage changes in distance from the object hence, because of geometrical relationships, on the size of the cortical image. Moving from a distance of two feet to one means a very drastic change in the population of cells which are being activated or inhibited as the case might be. The problem is aggravated, not lessened if we take fixations into account. Typically the eye will fixate regions of interest e.g. one of the four corners, with that corner at the foveal projection the rest of the screen is painted on an even further away piece of cortex; or text, which is usually entered at the bottom of the screen. If the eye wanders over the screen to get the complete picture things will really get worse, not only the image is all over the cortex, but the same piece of cortex is not even getting the same part of the picture all the time.

Any scheme that attempted to collect all of the elementary features present in each part of cortex at each time to then link the relevant ones together to form objects in an ascending hierarchy would face incredible difficulties.

On the other hand consider ISAN's image understanding method-- as long as some internal state activates a control function then it does not matter what part of the image arrives where, it will

arrive with high gain if it belongs to the object of interest. Furthermore different pieces of the image in different parts of visual cortex do not need to become integrated either there or somewhere else-- all of the image belonging to the object of interest is enhanced. Thus we would predict, post facto unfortunately but that does not lessen this data driven requirement for any model, that lesions or cross cuts or pieces of gold foil in visual cortex () should not impare visual performance. The reason ISAN performs in this way is because the lateral inhibitory function is local, even though its effects, in terms of selective image amplification are global.

From this point of view, even though ISAN has nothing to do with holography, ISAN resembles a hologram-- here too local recordings of intensity and phase physically unconnected to each other have global results in terms of image reconstruction.

5.3 Comments on learning rule

Even though the cortical model we have described works very well in terms of reproducing experimental results with a structure that is very close to known neuroanatomy there are some reasons to be unsatisfied with the learning rule as described. There are in fact situations in some animals where visual learning appers to be as that in the model e.g. imprinting in birds. During a

brief period, shortly after hatching, some birds will imprint on any nearby object that moves-- sometime with comical results. Ducklings can imprint on a hen which then greatly agitates when they jump in the water, on striped balls etc. Imprinting takes place extremely fast and requires no reinforcement. Not everyone agrees that imprinting is a case of learning.

In general it would appear that learning takes place more slowly and that some form of reinforcement is needed. To be simplistic about it this would assure that a causal relationship between a desirable event and an average of the event(s) that preceded it is recorded. Unreinforced events would not be recorded or learned thus freeing the animal from using up memory space with the irrelevant. There is a vast literature on learning experiments and learning theory (see Dickinson for a good review on the subject). There is reason to believe that if a simple learning rule could be identified that would account for the phenomena of classical and instrumental conditioning it could greatly improve brain models in general and vision models in particular (Klopf), such as the one described herein, that depend for their performance on the learning of domain specific images.

To strike a doubtful note it must be pointed out that learning theory deals with the simplest, structurally, systems and completely bypasses the complexities of those mechanisms that

deal with image processing-- another way of saying this is that the best of rules still requires the proper architecture for the system to work.

The perceptive reader must have noticed that this model asserts that a given object would be seen, just as it is, regardless of the actual shapes of receptive fields in visual cortex. It would of course take a bit longer, the object would not stand out from the background in a way somewhat related to how interesting it is, it would be much more easily masked by noise etc. The almost iconic recording of experience in the shape of the receptive fields in visual cortex is necessary to achieve selective image amplification in subsequent situations with resulting fundamental advantages in terms of speed, object seeking without having to start the analysis from the atoms of vision, and superior noise resistance.

ISAN is astonishingly good in separating objects from background and noise regardless of position if it knows what it's looking for. The cortical model makes the assumption that, if a new control function has to be learned, it will be locked in at the center of vision. Furthermore a startup set of control functions, genetically built in, has to be present. While early experience unquestionably adds to this initial set this bootstrap is, in our opinion, responsible for the so called Gestalt grouping principles

which seem to preexist individual experience. If there is merit in these considerations and in ISAN's approach it should now be possible to identify these control functions, whether they are embedded in the hardware or dynamically asserted by the efferent system that enforces them, or both.

6. SOME EXPERIMENTS WITH THE MODELS.

We run a large set of experiments with ISAN and a few with CORT. The following considerations guided the selection of the types of experiments to be done.

- 1) An image understanding system, conceived as an image seeking mechanism, needs to enhance meaningful images directly, so that neural activity produced by them will clearly stand out in the brain. ;We don't mean that a hot spot is produced, the whole image is always carried by a complete code;.

- 2) Subjectively, when we see an object, we don't see irrelevant contrast, but we often see relevant contrast quite clearly even though it is embedded in noise or actually absent. We experience no difficulties with image motion, i.e. translations, rotations, size changes, spectral variations etc. Irrelevant detail is not eliminated however-- it is only a posteriori that relevance is established, that is why a complete code is needed.

In summary it could be said that image seeking more often than not, requires not seeing what's there, seeing what's not really there and the ability to cope with image variations.

3) An image understanding system needs a great deal of start up knowledge that has phylogenetic and ontogenetic meaning. All that follows will be processed, interpreted and stored in this light. A continuum is thus required in the memory mechanism beginning with genetically hard wired structures (phylogenetic memory), and ending with structures firmly wired during development and less firmly later on in life (ontogenetic memory). The principal feature of the memory mechanism we develop is that memories are not just on the way to outputs, but are control functions that reflect back to the input. What we know, at all times and unavoidably, affects how we see, not just how we interpret, and the same is true of CORT. Other, though less interesting, features are that it accounts for column formation, is extremely fast, efficient and potentially resistant to damage.

6.1 EXPERIMENTS WITH ISAN

- a) forward and backward functions
- b) translation constancy
- c) size constancy
- d) orientation constancy

- e) noise suppression
- f) subjective edges
- g) sum of functions
- h) different functions at each stage
- i) holistic functions

t 1.1 Forward and backward functions.

The names forward function and backward function both refer to lateral interaction functions. We use the adjective backward to distinguish functions that were arrived at by specifying an input and a desired output and then solving the equations backward to find the set of coefficients that would thus perform. On the other hand we call forward those inhibitory functions that were developed using ISAN's efferent loop or by making them up using known types of receptive field shapes known to occur in visual cortex. We especially concentrated on the types of receptive fields that we have demonstrated to be produced by early experience. Thus the choice of three vertical or horizontal bars is far from coincidental-- they are the exact images that we used in our developmental experiments so that we also know exactly what types of functions will be generated. Fig. 4 shows the performance of one of the backward functions geared toward selective amplification for a vertical zebra.

In this, and all subsequent pictures representing ISAN's runs, the four panels represent image, output of stage one, stage two and stage three respectively. To avoid lengthy runs our SUN station is unfortunately based on a non-parallel architecture; the networks are 27 cells wide only. Activity levels are denoted using pseudocolors from dark blue to red to mean 0 to 255.

The function for vertical zebra performs quite well-- notice that vertical edges in the kite are not amplified, but that the vertical zebra is. We have not abandoned the hope that there should be a mathematical method to identify functions that are capable of inhuman precision in selecting for the desired output; in fact we do have functions that out-perform the human observer; however the least square approximation method we used is clearly not the one. Simply letting ISAN recycle the image to develop its own function always worked much better and using criteria gleaned from our experiments on development was best of all. We quickly discovered that the reason for this result is that there are sets of biologically plausible coefficients which, mathematically, produce an unstable system. In the visual pathway however only a limited number of cycles is possible because of the limited number of layers-- the reaction is quenched before it gets out of hand and selective amplification is obtained ;radio amateurs might remember the super-regenerative radio as an example of harnessed instability or failing that reflect that

life itself is an unstable system. On the way to disaster however very many pleasant happenings take place. These functions are inaccessible to systems of simultaneous equations because they don't exist in the solution space. We are consulting with a mathematician on other possible methods.

6.1.2 Image translations-- or translation constancy.

ISAN has no cells that will fire to, that is recognize, an object or a feature regardless of position. What it has is selective amplification for that object regardless of position. The reason for this performance is, as already mentioned, that all cells have local and identical connectivity to their neighbors thus any part of the network performs the same type of selective amplification as any other. The only exception, unavoidably, is at the edges as there are no lateral connections coming in from the outside. Position of the object of interest, and all of its parts, is retained at all times and is denoted by those cell in the network that are more active. As repeatedly pointed out the efferent control function affects all cells in each layer, much as in radio in which all stages of amplification are tuned to the signal of interest. If the control function is a simple gaussian then contrast, that is the edges of any image would be enhanced irrespective of where the image is in the network. Similarly complex functions amplify a specific image more regardless of

translations on the x or y axis. Fig. 5 demonstrates that the vertical zebra is enhanced even though its position has changed. This rather straightforward consequence of complex lateral interactions is extremely important in terms of performance of the visual system-- knowing where things are, at all times and as precisely as possible, is, quite simply, vital ;see also Marr on this point.

6.1.3 Size changes.

There are really two types of size changes-- in one scenario ISAN has developed a control function based on an image, e.g. a zebra of a certain size, and is then presented with a larger one even though the spacing of the stripes remains the same. The question here is will it be seen as one region or as a combination of regions? In the language of Fourier analysis the high spatial frequencies of the two zebras are the same, but the low frequency component, that is how long the zebra is different. ISAN sees the longer zebra as one as is shown in Fig 6. Even though this is good, we see the same thing, it is the least interesting case. The more interesting one is when the high frequency component also changes which is what would happen if the zebra were further away or closer than it was when its image was first learned. In this case ISAN's size constancy ability is about 50% that is ISAN can tolerate 50% change in size before its response falls off. This

result is shown in Fig. 7. The indication here seems to be that full size constancy require learning more than one control function. From the point of view of the theory, the requirement that at least **some size constancy** should be present in the basic structure from the start to sustain evolutionary pressure is fulfilled. However, major questions remain unanswered-- when an object approaches from afar, the size of the retinal image **doubles** every time that the distance is halved-- the very unpleasant consequence of this law of optics is that the largest and most sudden changes will occur at close distance, just when **time is least available**. Notice also that most animals manipulate objects with their mouths, which would seem to compound the problem, and that, simply because of size ratios, smaller animals ;with correspondingly smaller brains¿, have to get closer. One can only conclude, as anyone with impaired vision can readily confirm, that making objects larger than normal is not a problem, in fact recognition is facilitated. We have no answer as to how this can be, only a suggestion to be investigated later and that is that once the the retino-geniculo-cortical loop has locked onto an oject it will stay tuned to it over large ranges of variations. It would make the visual system vulnerable to **safe** objects that transform into **dangerous** ones-- this also happens and is known as camouflage.

6.1.4 Changes in orientation.

Perceptual constancy in animals in the face of changes in orientation of the retinal image varies depending on the source of the change. Rotations around the z axis, such as those produced by tilting the observer head are quite well compensated for i.e. the world doesn't tilt. If the world tilts it's immediately sensed and there is no constancy for it ;try reading upside down text;. The former type of constancy requires vestibular and/or proprioceptive information. ISAN demonstrates, as expected, a small amount of rotational constancy around the z axis. Rotations around the y and or x axis fall in a completely different category as rotations around these axes would bring into view hidden parts of the image. These rotations must engage higher level mechanisms than those at present possessed by ISAN and were not investigated. We would guess that real animals with minimal brains would solve this problem by storing many views ;memory is cheap; of the object of interest. There is in fact some evidence that humans do this too for images that are in constant use such as letters of the alphabet.

6.1.5 Not seeing what's there-- or ignoring irrelevant edges.

This is in fact one of ISAN's capabilities that we studied the most. In a natural environment, e.g. under a tree, objects might be illuminated by speckled light, sunlight might fall on one half

of a white piece of paper, etc. Our vision is untroubled by such everyday occurrences. The question is do we have to reconstruct the page-- after a feature analysis of all that irrelevant contrast or do we avoid seeing it to begin with and see it only if we want to? ISAN demonstrates that large amounts of random intensity variations overimposed on the image can quite easily be ignored. In Fig 8 the vertical zebra emerges from noise, whose maxima and minima have the same range of contrast ratio of the vertical zebra, very well, in fact surrounded by a halo of inhibition. More recent results, see below, show performance that exceeds that of the human observer.

6.1.6 Seeing what's not there-- meaningful but nonexistent edges.

Subjective edges are edges which are clearly perceived even though there are no physical correlates in the image that could account for them. They are induced by parts of the image which can be quite distant, but are arranged in such a way as to delineate some small parts of a physically non present object. Typically, black circles have missing sections as if another object, e.g. an ellipse or a square was occluding them. See as an example Fig. 9.

What is so remarkable about subjective edges is that we don't just infer the presence of an occluding object, we actually see it as brighter than the background. Further the inside of the

subjective image is also and uniformly brighter than the background even though, physically, there is no difference in the intensity distribution. The square frames around the circles are not strictly necessary, but they do further enhance the subjective image. This phenomenon is one of the many arguments of Gestalt psychology to demonstrate the insufficiency of atomic analysis of the retinal image and in favor of global a priori mechanisms as the foundations of image understanding. ISAN can see the nonexistent square just as easily as it sees a real one-- Fig 10. Interestingly subjective edges are just as resistant to noise as real ones-- see Fig. 11 in which ISAN's performance is better than human!!

In ISAN also, the presence of the outer frames improves the result even though they are not strictly necessary. There is no need to infer first that there is a square and to enhance it secondarily, what is needed is an **expectation** of square, that is a control function that is selectively amplifying squares, in fact the same function that selectively amplifies real squares works for the subjective ones as well. On the surface this requirement seems to trivialize the result, it does not for two reasons. The first reason is that, even if a specific expectation has to be set up, either because we are primed, and we normally are, as to what we are **supposed** to see by the text (illusions) or because there are unambiguous squares around the **subjective one**,

this is the only neuronally based mechanism that we know of that can account in a simple way for such a global perceptual phenomenon. The second reason is that this result seems to provide a key, by referring to well known Gestalt perceptual phenomena, that might unlock a very important door. It's beyond this door that we might find more such global functions whose combined action accounts for perceptual phenomena that seem irreducible to the sum of the parts which are present in the visual array.

6.1.7 Sum of functions.

The lateral interactions in each of the three layers in ISAN are realized by a system of linear equations, thus the principle of superposition holds. This principle states that in a linear system, and to repeat ISAN is conceived to be linear, the result obtained by summing the results of two transformations are identical to summing the functions first and applying the transformation later. Fig. 12 demonstrates that this is indeed how ISAN behaves. The image of the horizontal and of the vertical zebra are enhanced while all others are not. The control function that produces this effect was produced by simply adding the control function for kite to the control function for zebra. Because the principle of superposition is such a fundamental property of linear systems, in fact it is the defining property,

we did not invest a very large amount of time reassuring ourselves that it works. We are however convinced that because of this property a number of control functions capable of accounting for Gestalt perceptual grouping principles could be present ab initio and/or acquired during developement. The existence of gestalt grouping principles is inferred from figural effects that cannot be accounted for by the physical properties of the retinal image, precisely ISAN's domain of action. Grouping principles deal with symmetry, repetition, subjective contrast ;which we have demonstrated;, continuity, etc. Grouping principles seem to be the necessary foundation upon which vision is built (Rock). ISAN's ability to see subjective edges suggests that other control functions of general significance might exist and that, in combination, they might help explain how grouping principles come about. Estabilishing the existence of such a set of functions is left for the future work. If this endeavor were to be successful it would go a long way toward building a link between two approaches to brain studies that as of now seem irreconcilable.

6.1.8 Using different control functions at different stages

In the original ISAN model the same control function is applied to stage one two and three. We have pointed out above that efferents to retina are more or less abundant dependent on the specie. The structure of origin of the efferent fibers can also

be different, e.g. collicular for the retina, cortical for the geniculate nucleus. The lateral geniculate nucleus also receives inputs from non visual structures, e.g. the pons via the pontine-geniculate tract. We wanted to see if better image seeking could be achieved by using different control functions at different levels rather than a sum of different control functions identically at all levels. The rationale for this tests is that on one hand it would more closely parallel the anatomical reality and on the other it would put less demands on any one stage. As a possibly helpful analogy think of a single large antenna receiving signals from many stations and then using a broad band amplifier to amplify them all before going to a cable from which individually tuned sets can select the station of interest. Instead of having a perfectly flat frequency gain characteristic for the amplifier it would be more advantageous to have peaks of higher amplification for stations known to be faint. Even non selective broad band preamplifiers are very useful in improving signal to noise ratios and are commonly used. We have run very few experiments to test this idea, one however is of special interest and that is the situation in which the two retinal stages do contrast enhancement and the global seek function is sent only to the lateral geniculate nucleus. This is the condition that would apply in animals that have little or no efferent control to the retina and strong efferent control to the lateral geniculate nucleus. When the control functions are

applied in this fashion ISAN seems able to cope with considerable more noise. We are still investigating this approach.

6.1.9 Holistic functions

Up to now we have discussed, and experimented with, functions that selectively amplify specific objects. This is ISAN's basic theoretical foundation, and its implementation is that of an Image Seeking Adaptive Network. Here and there we have suggested that a small number of such functions in combination, thanks to the superposition principle, could account for the perceptual organizing principles studied by the Gestalt psychologists. We further suggested that some of these functions must be at the foundations of image understanding in organisms because of their biological significance.

Quite possibly one the most important elements in the visual world of organisms is the perception of the terrain, the horizon and the sky ;THS from now on. We have argued this point elsewhere (10). These elements of the image are the background upon which everything else has taken place throu-out evolution for millions of years. This is far from an original idea, however it necessarily follows that the mechanism that is responsible for the most elementary features of image understanding, ISAN in our theory, needs to be capable not only of what has been discussed

above, but also of some ability to see terrain, sky and horizon because these are amongst the most primitive and elementary images that organisms encounters. These words are used deliberately to point out that "elementary" has a very different meaning, in our opinion at least, in biology than it has mathematically or in computer science. Biologically it is simple to produce antibodies and to recognize complex molecules with them!!

We have run a very large number of experiments to produce a function capable of selective THS amplification. In the end, glossing over all the dead avenues that is, we uncovered a very simple function which is based on the idea that no matter where one looks in the distance there is a vanishing point. Thus early experience and/or natural selection would produce cells with receptive fields embodying that property. The property is that of a-symmetry between top and bottom ;terrain-sky?. Also, from the vanishing point, things grow larger. A receptive field, function, was thus designed: a-symmetric around the center and larger at the bottom. We then tested it with a variety of images in which the terrain was simulated by lines converging toward the horizon-- Fig. 13, by horizontal lines more closely spaced toward the horizon-- Fig. 14, or by random dot also more closely spaced toward the horizon-- Fig. 15.

A small road was also signified by arranging a few dots in a curved pattern-- Fig. 16. Large amounts of noise do not deteriorate ISAN's performance significantly and ISAN is capable of locating THS and the road quite well.

What is so exciting about these experiments is that regardless of the fact that they constitute a mere beginning in the direction of general vision, they compellingly indicate that this type of Gestalt functions exist, and that just a few would enable an organism to move, fly or whatever in a natural environment [given that however imperfect just one does THS], with no more than ISAN's basic structure. It hardly needs to be mentioned that, even if one disagrees with ISAN's theory, THS detection has to be one of the simplest and earlier mechanisms to evolve. A rival theory would have to provide a simpler and/or faster mechanism than ISAN's to be viable.

6.2 EXPERIMENTS WITH CORT

6.2.1 General comments.

Even though we invested a very large amount of time and effort in the design and programming of CORT, because of the very many aspects of cortical structure and function that the model has to account for, we have relatively few experiments with the finished

product. Very many experiments done to test unsuccessful hypothesis will simply remain invisible except for the remark that unsuccessful hypothesis were uniformly characterized by lack of simplicity and/or critically sensitive parameters. We found the evolutionary approach detailed above to be a very powerful heuristic in this endeavor. The final version of the model is structurally simple, has no critical parameters and succeeds in accounting for our experimental data on cortical plasticity and some of the major known features of cortical organization. As previously stated the model addresses itself only to the efferent module to the lateral geniculate nucleus. However because of our design philosophy, which has been to remain as close to the real system as possible and not just to build an ad hoc structure to explain our results, it also accounts for column formation, memory utilization and also the remarkable resistance to damage displayed by cortex. Most importantly the model also accounts for the origin of efferent control functions.

In the following paragraphs we will briefly describe some of the experiments that address what we consider critical and necessary elements of cortical function. ;Many vertebrates do not have a cortex in which case we would predict that similar functions would be performed, almost certainly less well however, by the optic tectum;.

a) developement of efferent control functions

by the cortico-geniculate loop

c) column formation

d) memory utilization

6.2.2 Developement of efferent control functions.

The idea that efferent control functions could arise by simply letting CORT and ISAN recirculate unknown images in the cortico-geniculate loop arose when we finally realized that there is a circularity in the argument that developmental plasticity is needed so that the animal learns to see. How can the animal learn to see specific objects unless it can see them to begin with? The circularity does not disappear by assuming that the feature detectors are built in-- in that case, why the incredible level of plasticity? Clearly a functional structure must preexist any experience and plasticity is critical. CORT needs some start up functions-- they are embedded in the center surround organization of the retina and in the initial structure of a few of its cortical receptive fields which we made elongated. This situation mirrors what is known to be present in cats at birth. Given these few seeds to start the loop ;the equivalent of thermal noise starting an oscillator; ISAN will capture an unknown object and CORT will learn the control function. Time is the price that needs to be paid because activity needs to be recirculated at least seven or eight times. Next time however the reaction is

much faster. Notice that in this conception the nature or number of the seeds is not important-- even if CORT is started with only horizontally elongated receptive fields a square would still be seen as a square and not as two horizontal lines-- there is data to this effect that could only be explained by this mechanism .

Selective image amplification under control of functions developed in this fashion is feasible. In general we found this to be an excellent way to develop control functions, in fact superior to the mathematical methods we investigated. A most important consideration needs to be stressed again at this point. We are convinced that real vision systems, as opposed to artificial ones have genetically built in start up functions which selectively amplify objects of interest to the specie, to these others are added during developement of special interest to the individual. While the ratio between these two sets of functions might vary in different species it is upon this endowment that all future perception will rest or fall. Further we feel that special functions that account for the so called Gestalt grouping principles are part of the first set. Finally it is quite probable, given the excellent vision exhibited by species without cortex e.g. birds, that these functions are implemented by phylogenetically ancient sub-cortical systems, such as the reticular formation, which are known to modulate lateral interactions in the input pathway. We intend to investigate this possibility in the future.

6.2.3 Column formation

Hubel and Wiesel conception of the functional significance of columnar organization in the cortex is that columns of different types, e.g. orientation, ocularity, etc. are needed to map the many dimensions of visual image's features (orientation, ocularity, etc.) onto cortex which is, because of limited thickness, to be considered either two dimensional or at most quasi-tridimensional. On this surface, much as in a quilt, the patterns of the different systems merge and intersect. The height of any column, equal to the cortical thickness, contains cells that are sensitive to small variations of the same dimension, e.g. position in a vertical orientation column. Because of their properties these cells have been named simple. There are also cells, within a column, that are functionally less sensitive to these variations, e.g. they retain orientation selectivity, but respond to a range of positions. Because functionally these cells seem to integrate the outputs of the simple ones they have been named complex. Hypercomplex cells have also been described. An impressive body of empirical evidence supports the above ideas. Disagreement, exists however as to the origin of these classes of cells (parallel versus serial, inborn versus developmentally acquired (3,21)), as to their meaning (line detectors versus spatial frequency detection (1)) and as to their actual feature

selectivity, i.e. these cells have been shown to also code for color, brightness, etc. (8). In our opinion a cortical model needs, to be meaningful, to take all these facts into account. However a model needs to do more than just account or be able to reproduce the data. A critical feature of model building has to do with what happens next. All of the models we know of, envisage visual cortex cells as feature detectors in a hierarchy at the top of which there exists a tassellation of values ;firing rates; indicating the goodness of match between each feature and each part of the image. It's not clear how large this mosaic would be, possibly down to just one element which would then be the proverbial grand mother cell-- however the original image would be lost!. In the model presented here a complete code is maintained at all times, thus what is passed on, in the case of an edge for example, is not a goodness of fit to an edge detector, but the actual edge enhanced or not depending on its belonging or not to a relevant image. Further, image integrity is preserved, in a global sense, by the efferent system. Because of these considerations we envisage the function of the columns to be quite different from that of a separator of image dimensions. In our conception columns originate as a byproduct of lateral inhibition whose purpose it is to prevent adaptation of other cells whithin its perimeter of influence to the image of the originator. Thus while in the end there will be columns, cortex is primarily where memory is i.e. the repository of the efferent

6.2.4 Memory utilization

As already mentioned we view cortex, primarily at least, as the structure where memory, in the form of control functions, resides. It would not be deleterious to our theory, in fact we argued that way a few paragraphs back, if memory, i.e. cortex were to be hierarckically organized with simpler memories in area 17 and more complex ones somewhere else, however because of the resistance to damage that cortex has, parallelism must be as massive as feasible. Therefore we believe, and the model supports this interpretation, that control functions for whole images are stored even in area 17 of visual cortex, which is the first structure after the thalamus. We reject the commonly held interpretation that at this site all that goes on is a decomposition of the image into elementary features to be recombined further on in the processing chain. This would constitute a giant step backward, functionally, and would require that, after some geniculo-cortical structures appear by random variation vision would continue to be served by the tectum while one waits for random variation to produce all those other structures which are needed to recombine the elementary features into meaningful objects. This line of reasoning grossly violates the evolutionary principle stated at the beginning of this section and that is that each variation needs to provide a selective advantage in and of itself to be selected for. In our

opinion the generally recognized idea that cortex is necessary for finer discriminations is, of course, true, but not because there are finer analyzers in the cortex ;after all the limiting factor is the retina; but because cortex provides more memory and better memory partitioning (25). In this way each area can specialize for certain types of knowledge and these parallel systems can then in parallel and simultaneously select for their own image of interest. This type of organization accounts for speed which is of the essence, but also for the fact that quite often ablation of secondary visual cortex produces only temporary deficits-- primary visual cortex will suffice-- things are just much better with more memory. There are some interesting analogies with computers with regard to available memory and graphic's resolution we will not dwell on this subject however except to bring it to the attention of the computer enthusiasts. We already pointed out that when information is presented in parallel to the many entities in a plastic network two problems are encountered-- one has to do with preventing all networks from being modified by the experience. Lateral inhibition combined with presynaptic inhibition of the afferents easily realizes a winner take all function so that, whithin a certain area only few cells are modified by the experience and the great majority remains unaffected. The other problem has to do with the requirement that those cells that manage to shut down all others in a wide area and will be modified by the experience, must be

uncommitted cells, that is cells that have not been previously tuned. This requirement makes sense logically and is also enforced by the data which shows that tuned cells do not retune(21). This proved to be a very thorny problem as there is no way to know **a priori** if a cell is not responding because it is not tuned to that image and uncommitted or is not responding because it is committed to a different image. A solution was found by allowing the lateral inhibitory stellate cells to be plastic also, in that way lateral inhibitory circuits that have been previously activated many times before are stronger so that, all else being equal, when lateral inhibition is on in an area tuned cells are inhibited more than untuned one. In this way the winner take all mechanism always picks an uncommitted cell whenever the image is new. An interesting outcome of this rather simple circuitry is that memories are stored more redundantly at the start and less and less redundantly as the number of uncommitted cells decreases. This feature of the model matches clinical data which shows that early memories are more resistant to brain damage.

7. CONCLUSIONS

The neuroscience community has been extremely productive in the field of vision research at all levels. Image understanding, by

its very nature, involves very many disciplines and is considered to be the interdisciplinary field par excellence. In many instances, however, generally important results have remained ensconced in their own domain either because they are old, pertain to a different animal, or seem simply unrelatable. As a result a general theory of vision supported by mechanisms that have been tested by computer simulation is still missing.

In these studies, and in the theoretical framework that we develop from them, we make a first attempt at relating our results on plasticity with modern learning theory, with Gestalt'conceptions on visual perception, with functional and structural anatomy of the retino-geniculo-cortical system, and finally with efferent control systems. We also provide mechanisms to support the theory developed and computer models as a prove that the mechanism proposed perform as expected.

In the course of these studies we developed a paradigm capable of producing **unique experiences** i.e. experiences whose attributes could not occur in the everyday experience of an animal without outside intervention. Typically this implied a danger image presented to one eye only and, following the appropriate behavior by one leg, presentation of a different safe sign to the other eye only. In other situations we simply alternated different patterns with varying delays from one eye to the other. These experiences produce powerful plastic phenomena and tune cells to response patterns that are also unique in that they do not appear

in animals that did not have this experience. We explain these **supernormal plasticity** by making appeal not just to development (other experiences during development are not as effective), but also to the well known phenomena of **sensory preconditioning and superconditioning** described and quantified by learning theorists. In other words the reason our paradigm works so well is because as one image follows the other we have sensory preconditioning; however because the first image sets up an expectation which is then violated by the second image, the surprise effect or the informative value of the second image is abnormally high which produces superconditioning. We conclude that, using these principles, it should be possible to design plasticity inducing experiences even in adults, which would have very important applications for psychiatry and adult education. In any case the design of future experiments on neural plasticity would benefit from taking modern learning theory into account. We incorporated **Gestalt thinking** in designing experiments and theory because of the nature of the plastic phenomena we observed and that is the plastic changes were always of a global nature because all the dimensions of the experience were retained. There are new opportunities, and problems, raised by this line of thinking. The opportunities lie in the direction of establishing a framework to begin to understand how modulation of local phenomena leads to global results; well known Gestalt perceptual phenomena could provide powerful suggestions as to the nature of

the imbedded functions now that we have a mechanism to enforce them. The difficulties are on the empirical front-- if testing for one dimension of sensitivity, e.g. orientation, or possibly two or three, i.e. orientation, position and length is not enough then the time required to analyze a single cell would increase out of bounds, enormously aggravating the difficulties of data gathering which are already substantive for this type of research. Even more difficult would be to take efferent control into account when mapping receptive fields of single cells. It would not be sufficient to just activate it as has been done up to now, but it would be necessary to activate it meaningfully. On the plus side experimentalists are incredibly resourceful.

In the anatomical studies, still in progress, we have used horse radish peroxidase to trace connections, Golgi-Cox to study changes in dendritic trees and a silver stain by Cajal to study dendritic bundles in those areas where we know, from single cell recordings, plasticity has occurred. As mentioned the peroxidase studies, in which we do tridimensional reconstruction using a SUN computer, show that the primary connectivity is cortico-thalamic, and that is the source of our emphasis on the cortico-geniculate loop in the theory and model. There are cortico-cortico connections, of course, but they seem relatively few in numbers. We have also shown that dendritic trees have more branches in the cortical area where plasticity has occurred, i.e. sensory-motor cortex of the trained leg, than in the same area in the other

control functions. One extremely important aspect of the experiments with CORT then was to ascertain that it would not only account for receptive field formation in a way that mirrored our results, but also that the learned modifications would in fact be capable of generating an output to ISAN that would selectively amplify the image that had caused it in the first place.

The experiments done to date with CORT show that columns do indeed form in response to vertical zebras and they resemble the receptive fields of real nerve cells. New images with features that cannot be reduced to this experience produce more columns without affecting either the columns previously formed or the receptive fields of the cells contained in them. The receptive field formed by the vertical and horizontal zebra exhibit more often just one excitatory elongated region with one or two inhibitory flanks, but there is also an occasional one with three excitatory regions. These results mirror our findings on cortical plasticity.

When an image that has already been learned is present in the visual array, e.g. a vertical zebra, the cells that have receptive fields tuned to that image will become active and will prevent other cells from learning it. Their activation will in turn excite the pyramidal cells connected to them whose output feeds back to the lateral geniculate nucleus layer thereby locking onto the image. If a new image is presented to the model

it will not find pretuned cells therefore the the output from the pyramidal cells will be very close to the input as acted upon by the basic contrast enhancement function. The output from the pyramidal cells modulates the inhibitory function in the geniculate layer enhancing the selectivity for that particular image and in about seven or eight cycles around the cortico-geniculate loop the image will be locked in. At this time the stellate cells are allowed to be plastic and the new control function is learned. We have shown above that this function is in fact capable to tune the geniculate layer for selective amplification. In our conception the receptive fields of stellate cells that form from experience are not much different from those that have been described by others, however on one hand we claim that the details of the receptive fields are just as important as other major (orientation) features because they do in fact determine the overall selectivity, on the other these receptive fields are not templates against which the image is matched and from which firing rates are sent on to the next stage to indicate the goodness of match-- rather they are the originators of the efferent function that will selectively enhance that image. Lets remember at this point one of the most important features of this process-- because of the complete code any detail or small imperfection of the object of interest, **is still preserved**, just as it would be in our subjective experiences.

hemisphere for the untrained leg. A most interesting finding has to do with dendritic bundles which are longer and contain more dendrites in the trained cortical area than the untrained one. The difference is very clear and visible just with the optical microscope. Dendritic bundles could be the structure responsible for cortical resource allocation-- if so it would become clear why some of the plastic changes are permanent. While synapses can change their strenght and even dendritic trees can shrink or expand, dendritic bundles cannot unbundle.

Finally we have organized all of this data and more from previous work of ours and other workers in a theory supported by a computer model that performs remarkably well. The computer implementation is particularly important in our case because the theory's central tenet is that the visual system of vertebrates operates on whole images. This idea is contrary to the general approach (with the exception of Gestalt psychology possibly) which advocates elementary feature analysis. We thus don't expect it to become a popular theory soon, but given that we can demonstrate that it works there's hope.

Regardless of its acceptance by the neuroscience community however the structure and the algorithm proposed do work and might be of interest in machine vision.

8. MILITARY SIGNIFICANCE

The research described here has potential on four fronts.

8.1 **First**-- the structures and algorithms implemented as models of the vertebrate visual pathway can form the foundations of an **image seeking device** which, if implemented in hardware, would be fully parallel and extremely fast. Because there are only four layers and images are **not decomposed into elementary features** it is conceivable that such a device could operate in just a few microseconds.

8.2 **Second**-- the physiological findings provide a blueprint of how to **induce maximum and fastest learning** by taking advantage of sensory preconditioning and superconditioning. These principles fundamentally alter the speed and magnitude of the learning process **whenever there is a double representation**, which is the case for most systems in the brain. Pilot's training could benefit by applications of this part of the theory.

8.3 **Third**-- if, as we have demonstrated, **what is subjectively enhanced in the visual array** critically depends on high level control functions that modulate what the low level selectively amplifies, then this research makes a precise statement as to the nature of the control functions. Pilot's performance could benefit from forms of training which are specifically geared toward task related control functions.

8.4 **Lastly**-- if we take Searle's Chinese-Room thought experiment seriously, and we do, it would seem that only by finally understanding the structural foundations of natural intelligence truly intelligent artificial systems could ever be built.

REFERENCES

1. DeValois, R.L., Albrecht, D.G., and Thorall, L.G. "Cortical cells: bar and edge detectors, or spatial frequency filters?" in **Frontiers in Visual Science**, S.V. Cool and E.L. Smith (Eds.), New York: Springer-Verlag, pp. 544-556, 1978.
2. Hayes, B. "A progress report on the fine art of turning literature into drivel," **Scientific American**, 249, 18-28, 1983.
3. Hubel, D.H. and Wiesel, T.N., "Receptive fields, binocular interactions and functional architecture in the cat's visual cortex," **J. Physiol.**, 160, 106-154, 1962.
4. Kolers, P.A., "Aspects of motion perception," Pergamon Press, New York, 1972.
5. Kuffler, S.W. "Discharge Patterns and Functional Organization of Mammalian Retina," **J. Neurophysiol.**, 16, 37-68, 1953.
6. Marr, D. **Vision**, San Francisco: W.H. Freeman and Company, 1982.
7. Palmer, L.A., Rosenquist, A.C. and Sprague, J.M. "Cortico-fugal systems in the cat: their structure and function" in Frigyesi, T., Rinvik, E. and Yahr, M.D. (Eds.), **Corticothalamic Projections and Sensory Motor Activities**, New York: Raven Press, pp. 491-522, 1972.
8. Pribram, K.H., Lassonde, M.C. and Ptito, M. "Classification of receptive field properties in cat visual cortex," **Exp. Brain Res.**, 43, 119-130, 1981.
9. Searle, J., "Minds, Brains, and Programs," **Behavioral and Brain Sciences** 3, pp. 417-418, 1980.
10. Spinelli, D.N. "A Trace of Memory," Arbib, M.A. and Hanson, A.R. (Eds.), **Vision, Brain and Cooperative Computation**, MIT Press, 1987.
11. Spinelli, D.N. "Texture viewing during development in kittens modifies the functional properties of single cells in visual cortex," **Soc. Neurosci. Abs.**, 3, p. 378, 1977.
12. Spinelli, D.N. "Receptive field organization of ganglion cells in the cat's retina," **Exp. Neurol.**, 19, 291-315, 1967.
13. Spinelli, D.N. and Jensen, F.E. "Plasticity: the mirror of experience," **Science**, 203, 75-78, 1979.

14. Spinelli, D.N. and Jensen, F.E. "Plasticity, experience and resource allocation in motor cortex and hypothalamus," in Woody, C.D. (Ed.), **Conditioning**, Plenum Pub. Co., 1982.

15. Spinelli, D.N., Pribram, K.H. and Bridgeman, B. "Visual receptive field organization of single units in the visual cortex of monkey," **Int. J. Neurosci.**, 1, 67-74, 1970.

16. Spinelli, D.N. and Barrett, T.W. "Visual receptive field organization of single units in the cat's visual cortex," **Exp. Neurol.**, 24, 76-98, 1969.

17. Spinelli, D.N., and Pribram, K.H. "Changes in visual recovery function and unit activity produced by frontal and temporal cortex stimulation," **Electroencephal. and Clin. Neurophysiol.**, 22, 1943-1949, 1967.

18. Spinelli, D.N., Pribram, K.H. and Weingarten, M. "Centrifugal optic nerve responses evoked by auditory and somatic stimulation," **Exp. Neurol.**, 303-319, 1965.

19. Spinelli, D.N. "Theory and Computer Implementation of an image Seeking Adaptive Network (ISAN): A New Approach to Image Understanding," **INTERIM REPORT: AFOSR-83-0207**, 1986.

20. Ratliff, F. **Mach Bands: Quantitative studies on Neural Networks in the Retina**, Holden Day, Inc., 1965.

21. Spinelli, D.N., H. V. B. Hirsh, R. W. Phelps, and J. Metzler. "Visual experience as a determinant of the response characteristics of cortical receptive field in cats," **Exp. Brain Res.**, 15:289-305, 1972.

22. Jensen, F. E. and D.N. Spinelli. "Early experience effect on dendritic bundles" **Society for Neuroscience Abstracts**, Vol. 4, 1979.

23. Feldman, J. A. and D. H. Ballard "Computing with Connections" **Technical Report TR-72**, University of Rochester, 1981

24. Spinelli, D.N. "OCCAM: A Computer Model for a Central Addressable Memory in the Central Nervous System" **Biology of Memory**, Academic Press Inc., New York, 1970

25. Spinelli, D.N. and F. E. Jensen "Brains, Man, and Crystals of Knowledge" **Machine Intelligence**, vol. 10, and D. Michie eds. (Ellis Hoewood), 1984.

Klopf, A. H. "Brain Function and Adaptive Systems" **Neuroscience Report AFCRL-72-0164**, Air Force Cambridge Research Laboratory, Bedford, Mass. 1972.

AD-A186 214

IMAGE UNDERSTANDING BY IMAGE-SEEKING ADAPTIVE NETWORKS
(ISAN)(U) MASSACHUSETTS UNIV AMHERST DEPT OF COMPUTER
AND INFORMATION S. D N SPINELLI 10 AUG 87

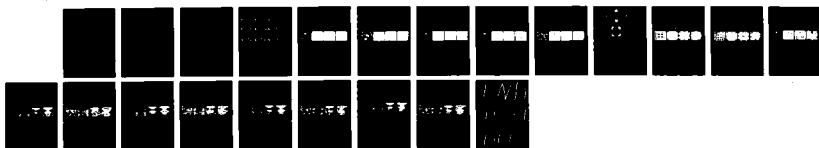
2/2

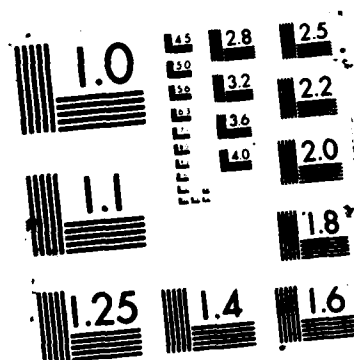
UNCLASSIFIED

AFOSR-TR-87-1206 #AFOSR-83-0207

F/G 17/11

ML





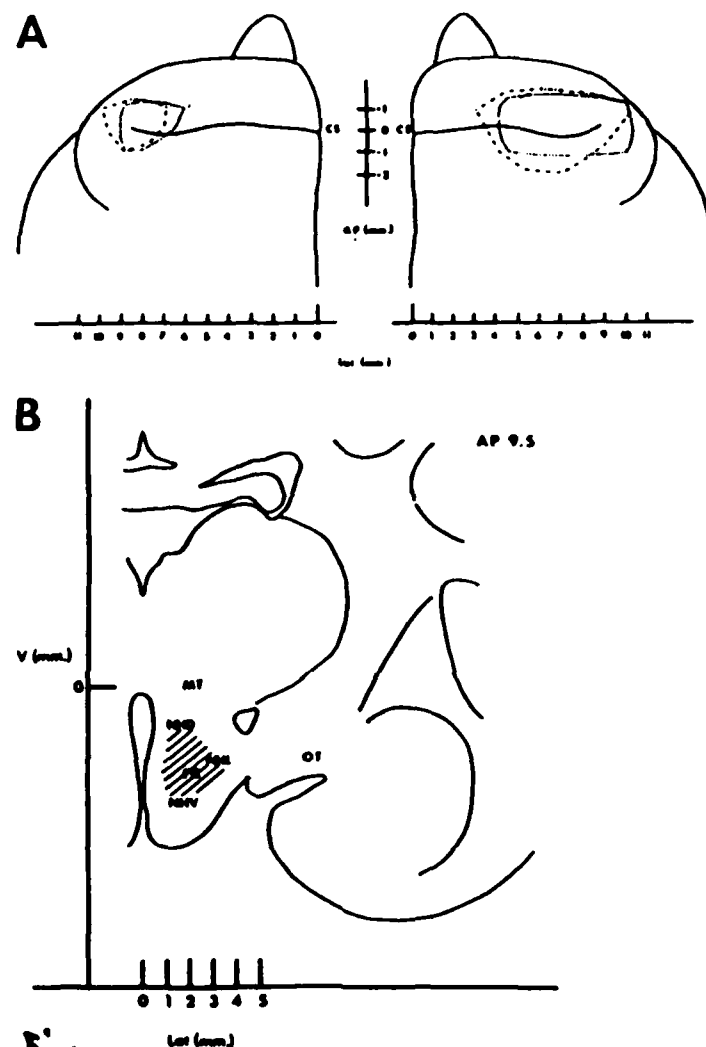


Fig 1

A: Varied line boundaries show the representational areas in motor cortex of three cats in which foreleg movement could be induced by cortical stimulation. Hemispheres contralateral to the untrained and the trained foreleg are shown. The area located to the control of movement in the trained foreleg was enlarged an average of 30 percent over that area concerned with movement of the untrained foreleg. B: Diagrammatic representation of a section taken through the cat's brain at AP: + 9.5. Striped area represents regions where cells were marked after recording and verified histologically. Ventral, lateral, and dorsal hypothalamic nuclei are labeled NVL, NHL, and NDL, respectively. Lateral and ventral stereotaxic axes are also shown. (Reprinted, with permission, from Spinelli and Jensen 1982.)

PERFORMANCE ON PLACEMENT TEST

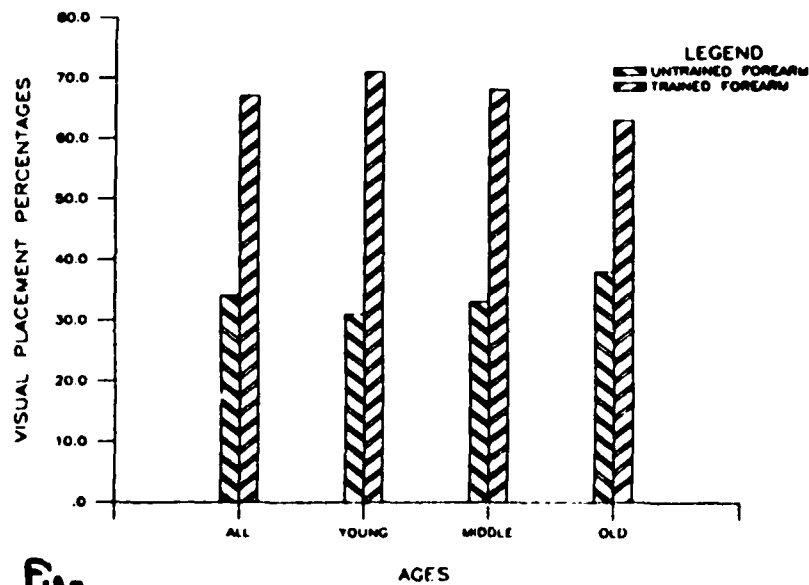


Fig 2

Results from visual placing test. The test consisted of simply holding the kitten in the air near a table edge and scoring which one of the forelegs touched the table first. The results clearly show that a kitten will make first contact with the trained leg in a statistically significant percentage of trials. In each pair of bars, the second bar represents the results for the trained leg.

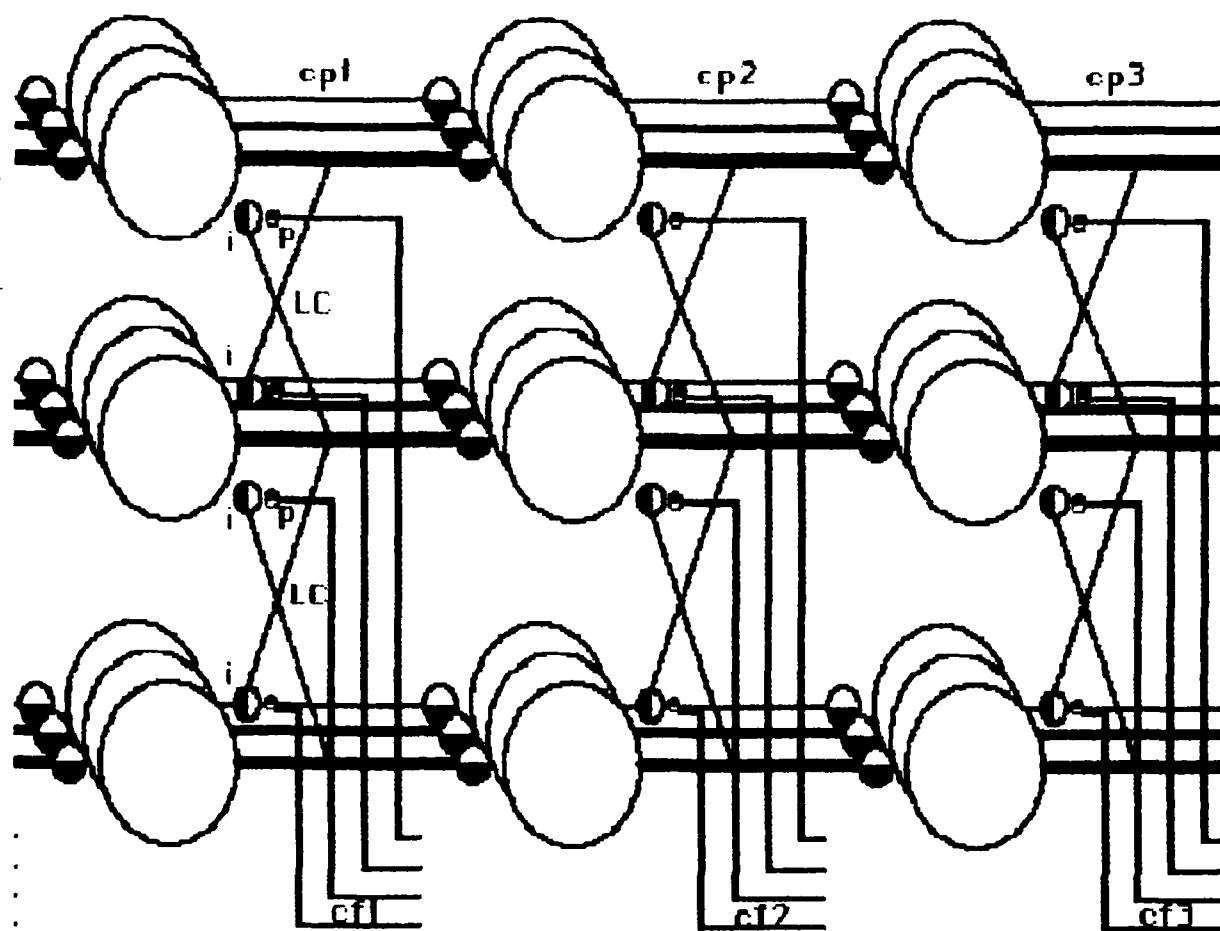


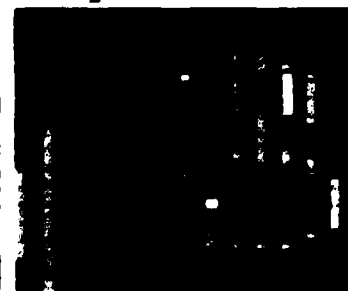
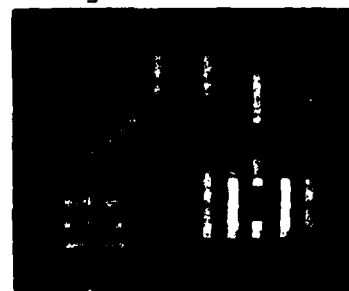
Fig. 3 ISAN's architecture showing three stages of a 3 by 3 two dimensional network. Cell bodies are denoted by circles, lateral connections explicitly labeled in the first layer only are LC. Connecting pathways from one layer to the next are noted by cp1, cp2 and cp3 respectively. Efferent control pathways from higher structures are labeled cf1, cf2 and cf3 to indicate that these functions need not be identical. Presynaptic boutons to cell bodies and for pre-synaptic inhibition are half full to indicate that sign reversal is allowed in the simulation.

Differential Image Amplification
 image
 image3
 noise=0.0
 amplify
 fzebrav.2
 noise=0.0

stage 1

stage 2

stage 3



000

pseudocolors

255

Fig. 4 In this figure and all that follow showing ISAN's runs the first panel represent the image, the second panel the output of the first stage of lateral interactions, the third panel output of the second stage and finally the fourth panel represent output of the third stage. Pseudocolors are used to indicate levels of activation of the unit elements. There are 27 by 27 units in each network. Dark blue signifies minimal activation and bright red maximal-- from 0 to 255. Images/background=2. Because there are only seven colors available in the printer that was available fine gradations in activity cannot be seen, however the difference in gain for the image sought e.g. the vertical zebra, is so large compared with the gain for the other images (horizontal zebra, kite, test bar) that the loss is immaterial. In this figure, and following ones unless otherwise stated, ISAN is seeking the vertical zebra using function fzebrav in all cases as shown by the legend above panel 2. Notice that while the vertical zebra is strongly amplified the vertical test bar and the vertical edges of the kite body are not.

Fig. 5 To demonstrate that the vertical zebra is still differentially amplified even though its position has changed. This result follows naturally because the lateral interactions are locally identical for each cell regardless of its position in the network (cells on the rim excepted). Random noise added to image = 66% of image amplitude maximum peak to peak.

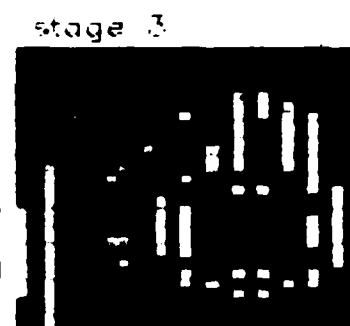
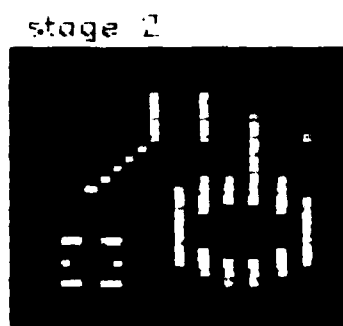
Differential Image
 Stage
 1
 Amplification
 stage
 1
 zebra1
 noise=0.0
 stage 1



000



pseudocolors



255

Fig. 6 To show ISAN's capability to selectively amplify vertical zebras even when they are **larger** both in size and number of bars. Notice again that **unimportant** vertical edges are either **de-amplified** or minimally noticed.

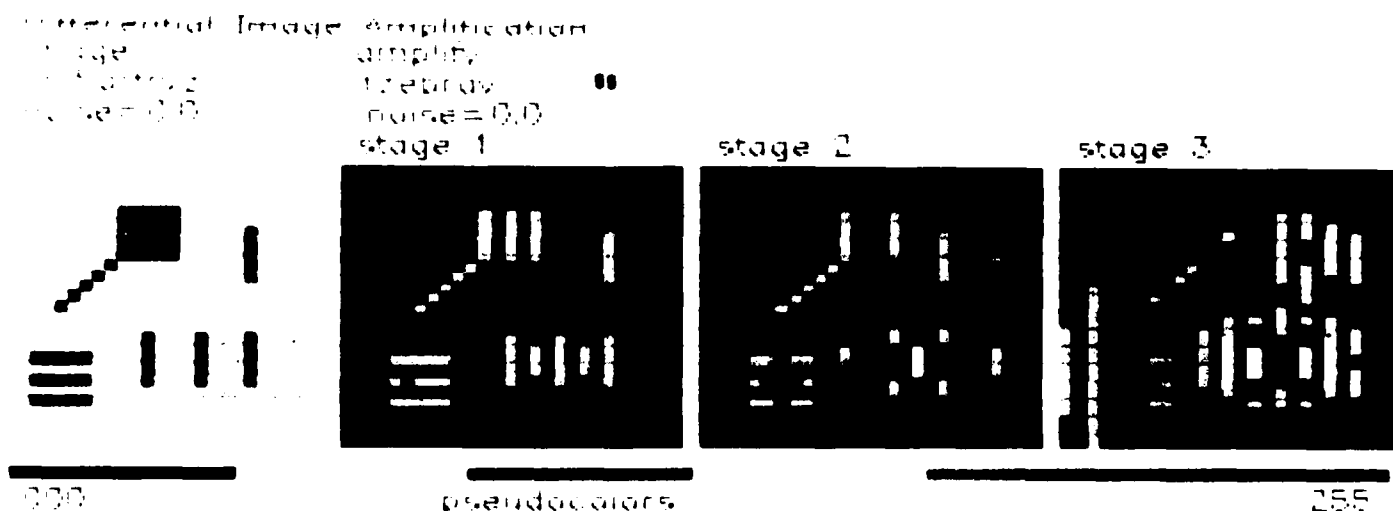


Fig. 7 Shows ISAN at the limit of its generalization ability while dealing with a vertical zebra in which the vertical bars are separated by double the normal distance. Notice that it would still enable the structure above to see (red level) that there is a **possible** vertical zebra out there and thus trigger other mechanisms. These could amount to nothing more than a number of **memorable** experiences with the object at different distances. Notice again that while ISAN is quite **elastic** with regard to the object of interest, it is still un-mindful of irrelevant vertical edges.

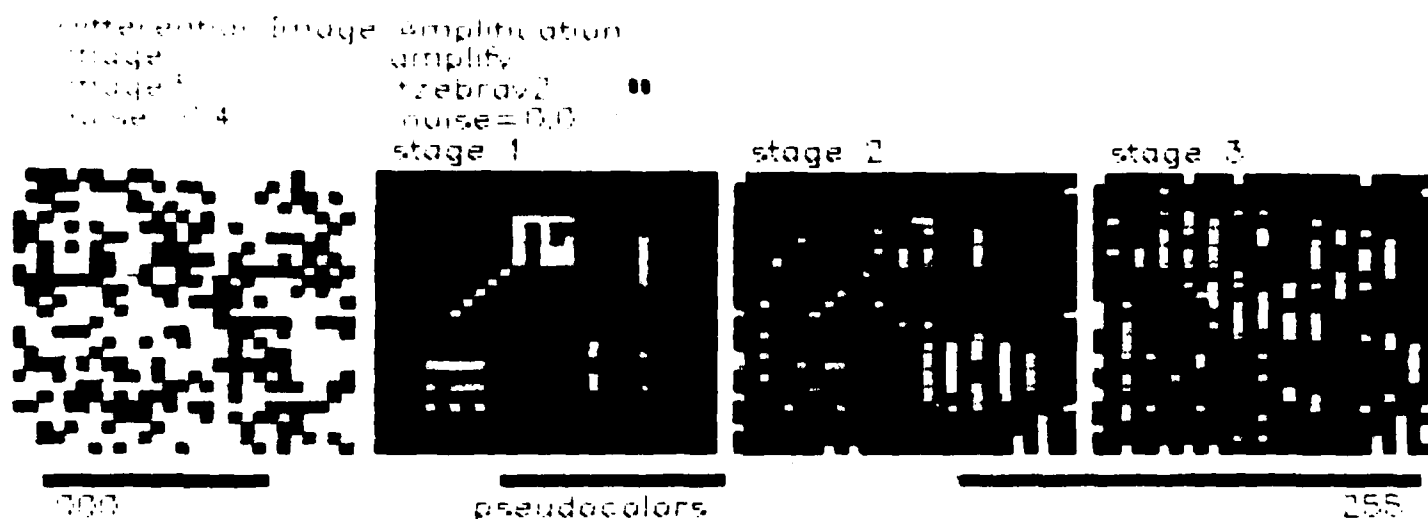


Fig. 8 Looking for vertical zebras in noise. Not seeing present but irrelevant contrast without the need for lengthy computations is, in our opinion, one of the most fundamental properties of natural vision systems. ISAN takes exactly the same time to selectively amplify (see) a vertical zebra with or without noise. In this respect it behaves very much like a radio-- noise is suppressed and the signal of interest is selectively enhanced. Random noise added to image = 66% of image amplitude maximum peak to peak.

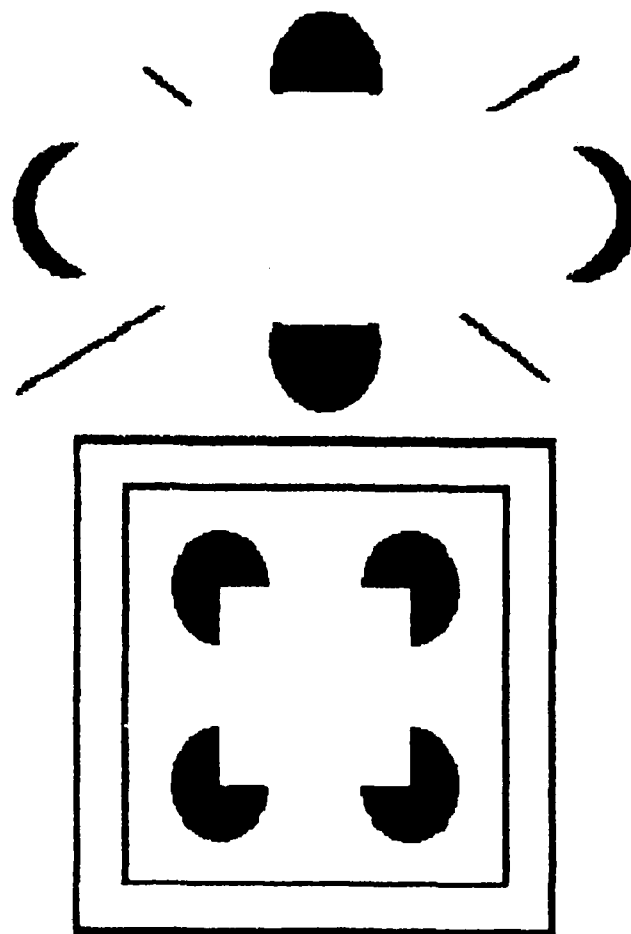


Fig. 9 Seeing non existent but relevant contrast without the need for lengthy computations is another most fundamental property of natural vision and it is something that we perform all the time and effortlessly, as this figure demonstrates. Even though there is no **brightness gradient** we see an ellipse and a square quite clearly. The ellipse has sharp elliptical edges and the square has sharp straight edges. Both objects look **uniformly brighter** than the background even though, to repeat, there is no **brightness** difference. Such images compellingly demonstrate that natural vision system have a **priory expectations** (organizing principles) that enable them to perform image understanding tasks that would be computationally intractable.

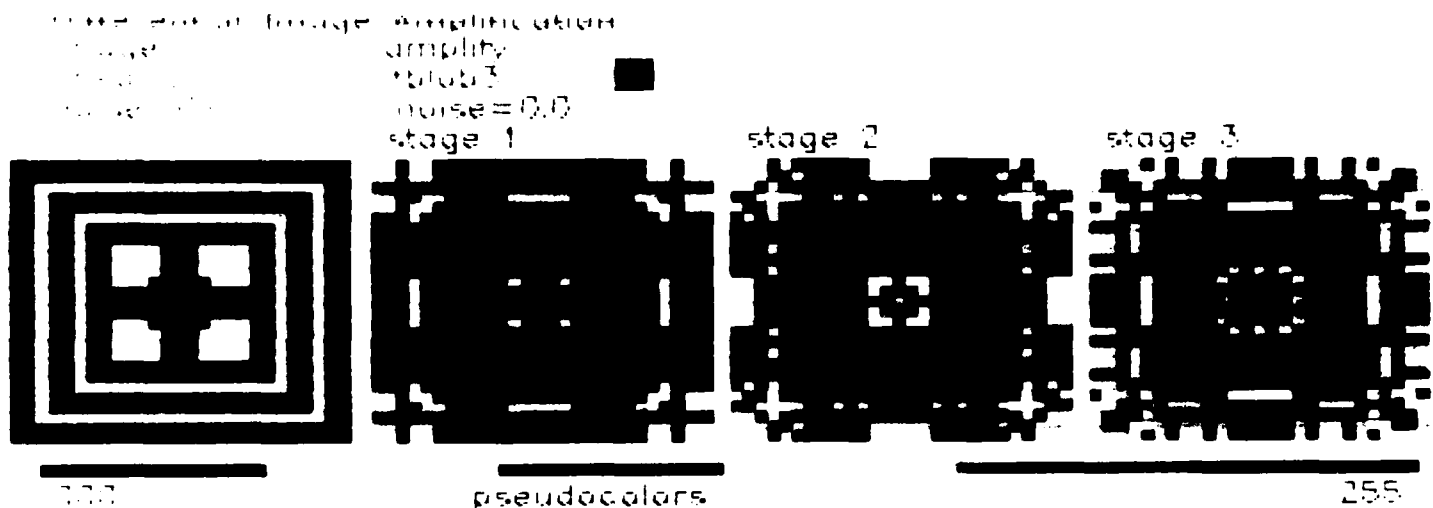


Fig. 10 ISAN's whole theory rests on the idea that the vision system is really an Image Seeking Adaptive Network(s). Zebras and other objects of interest can thus be seen directly. General vision however requires an immanent set of expectations that, while specific in and of themselves, are of general significance. ISAN's structure allows superposition of functions (see Fig. 12); thus it is possible to have more than one expectation. We envisage a small set of such functions to be responsible for the Gestalt organizing principles. This figure demonstrates that ISAN is perfectly capable of seeing the subjective square given an expectation of square. We can now build on the work of Gestalt psychologists to identify other primitive expectations necessary for general vision.

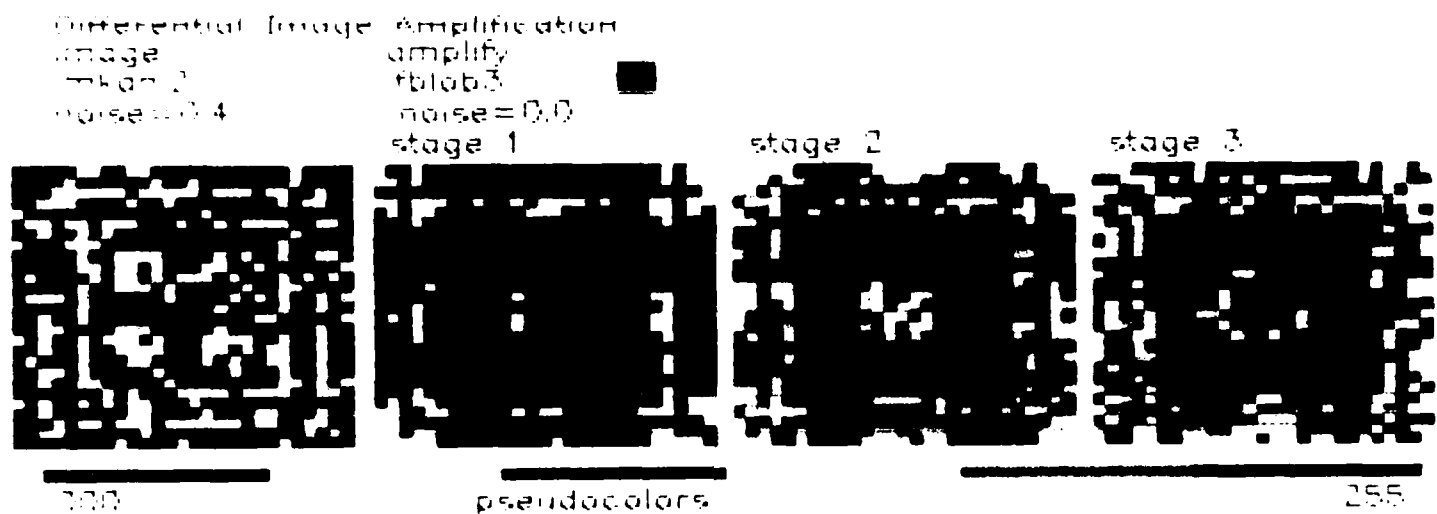
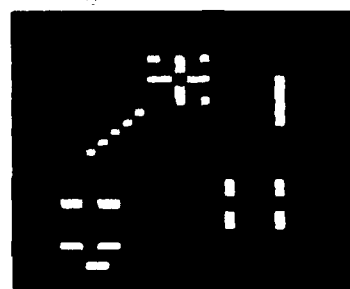


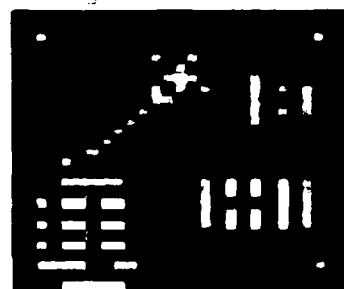
Fig. 11 Subjective square in noise. In this test ISAN still can see the subjective square even though humans, at least this one, can no longer do so. Recall that, with the exception of the edges, each cell in the network has the same connectivity and that differential amplification is enforced at each point- square is in the center purely for graphic purposes and to stay away from the edges.

Input image: A 16x16 pixel image of a zebra head and neck, with a black background. The zebra is white with black stripes. The image is labeled "Input image" and "16x16 pixels".

Stage 1



Stage 2



Stage 3



007

pseudocolors

255

Fig. 12 ISAN is a linear network that carries a complete code, that is **no information is lost**. Non linearities are left for the decision making mechanisms, e.g. those that decide to manipulate, chase or escape the **red image**. In linear mechanisms the principle of superposition holds-- thus it is possible to add two functions and look for two objects simultaneously **at no extra cost in time**. This property is extremely important, because **objects are made of objects in ISAN**. Here A function for vertical zebra has been added to one for horizontal zebra so that ISAN is seeking vertical and horizontal zebras simultaneously and indeed it finds them. Notice that even in this situation the vertical and horizontal edges of the kite are not amplified.

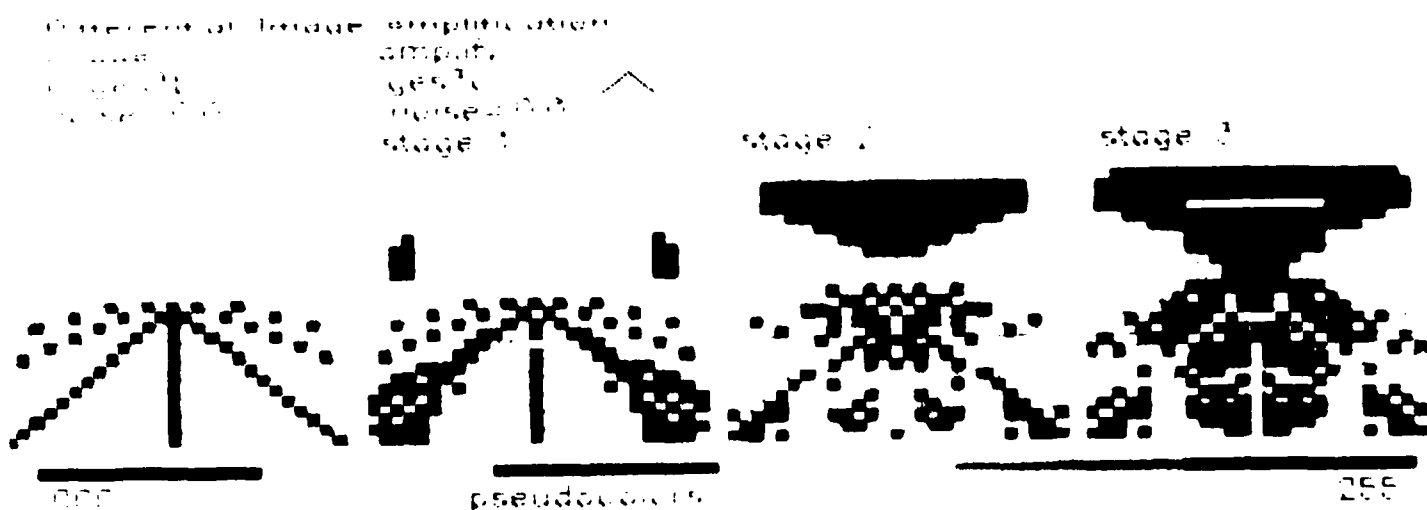


Fig. 13 In this figure an a-symmetric function, **ges3c**, is used to modulate ISAN's connectivity. A terrain, horizon and sky pattern (THS) is skeletally simulated in the image by a set of lines converging to the "horizon". Notice that the converging lines are "amplified"

54000

000000

କଟକରେ

pseudocolors.

755
756

Fig. 14 Ges3c seeking THS to which heavy noise has been added, horizon position has been changed and in which converging and parallel lines are used to skeletally simulate THS. Notice THS amplification and excellent noise suppression.

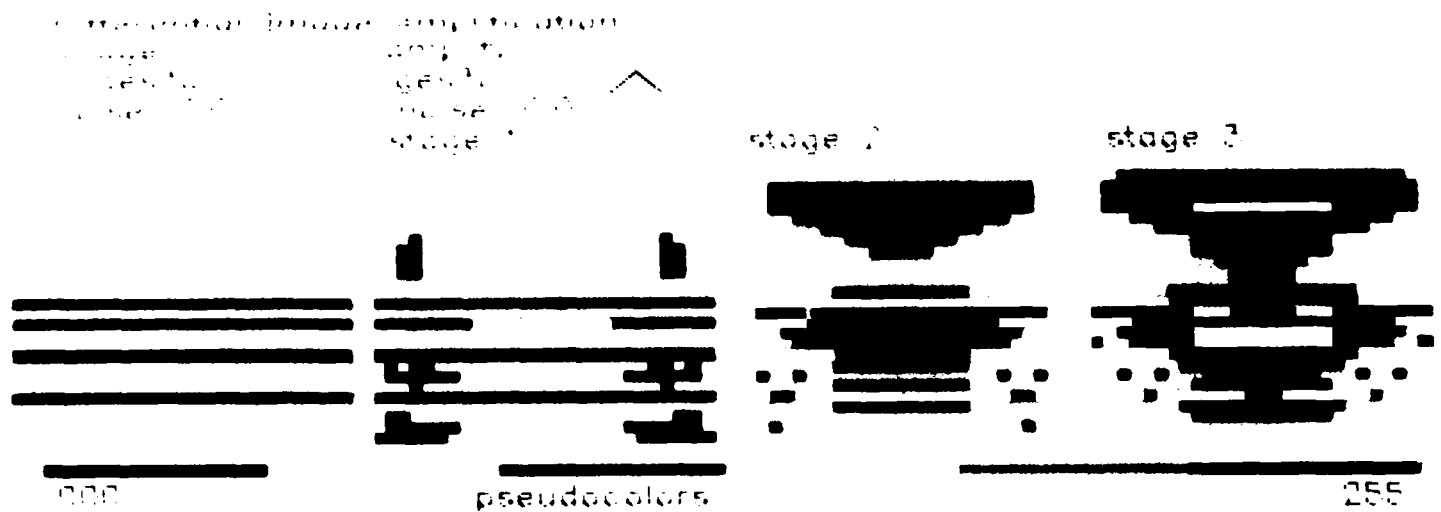


Fig. 15 The same function used in the previous experiment, **ges3c**, is used to **see** a THS skeletally simulated by horizontal lines. Notice that the horizontal lines are also amplified.

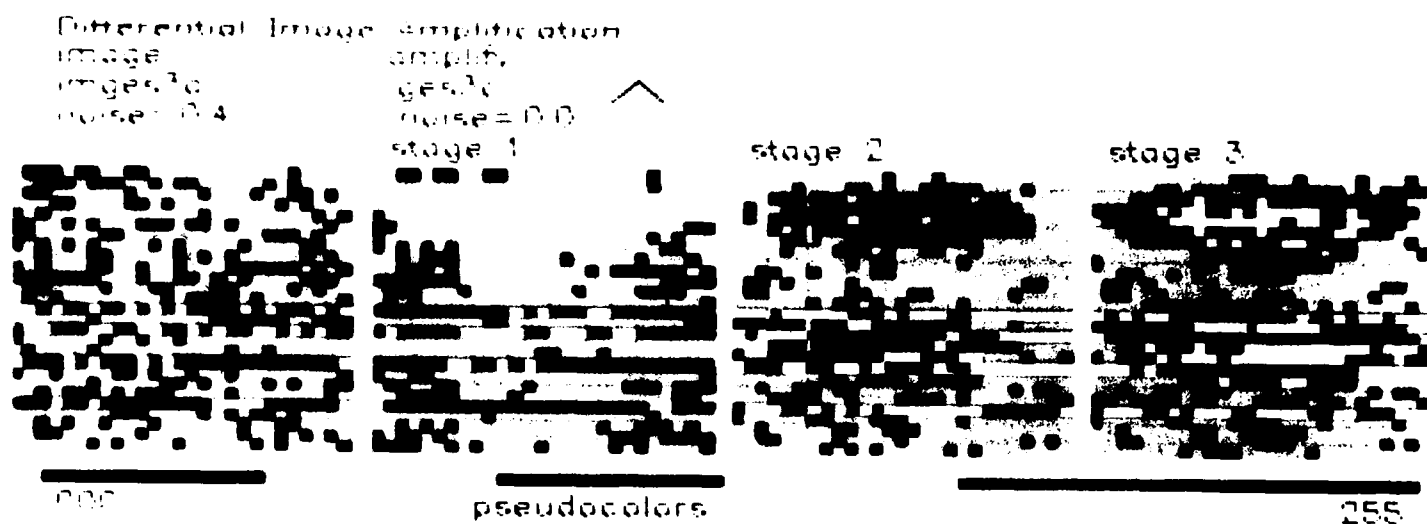


Fig. 16 Same details of Fig. 15, but heavy noise has been added. Again notice THS amplification and substantive noise suppression.

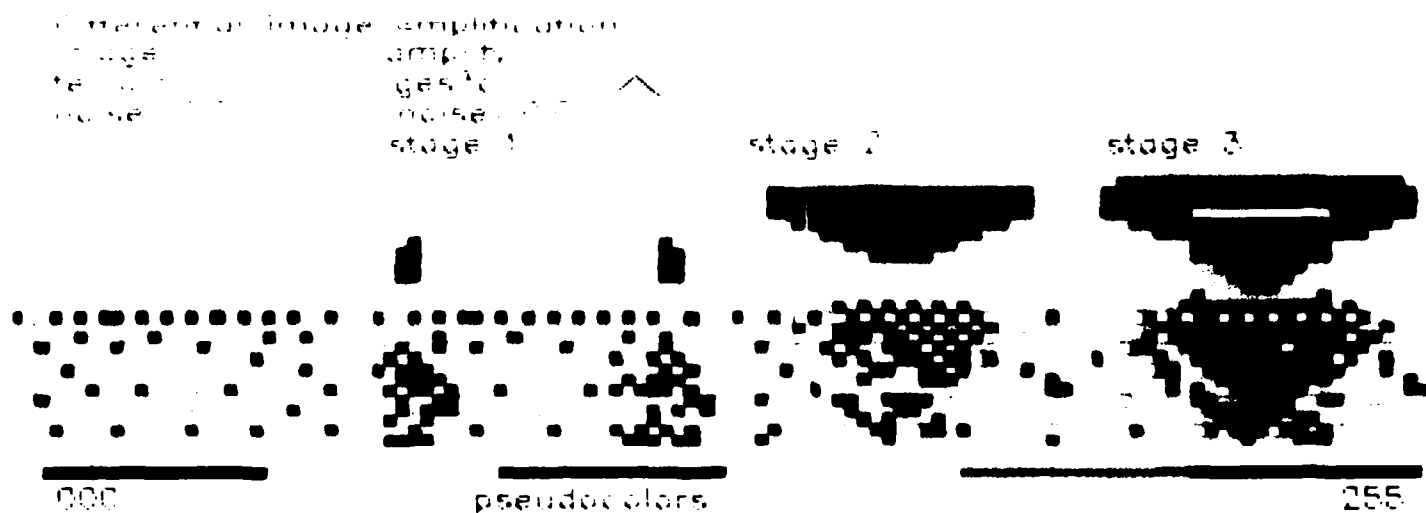


Fig. 17 **Ges3c** is used in seeking a THS skeletally represented by random dots of increasing density. Notice that this yet different THS is clearly seen.

Differential Image Simplification
 image
 to noise
 noise 0.4
 stage 1



Fig. 18 Same details as Fig. 17, but with noise. Again notice how irrelevant contrast is deamplified and relevant contrast selectively amplified.

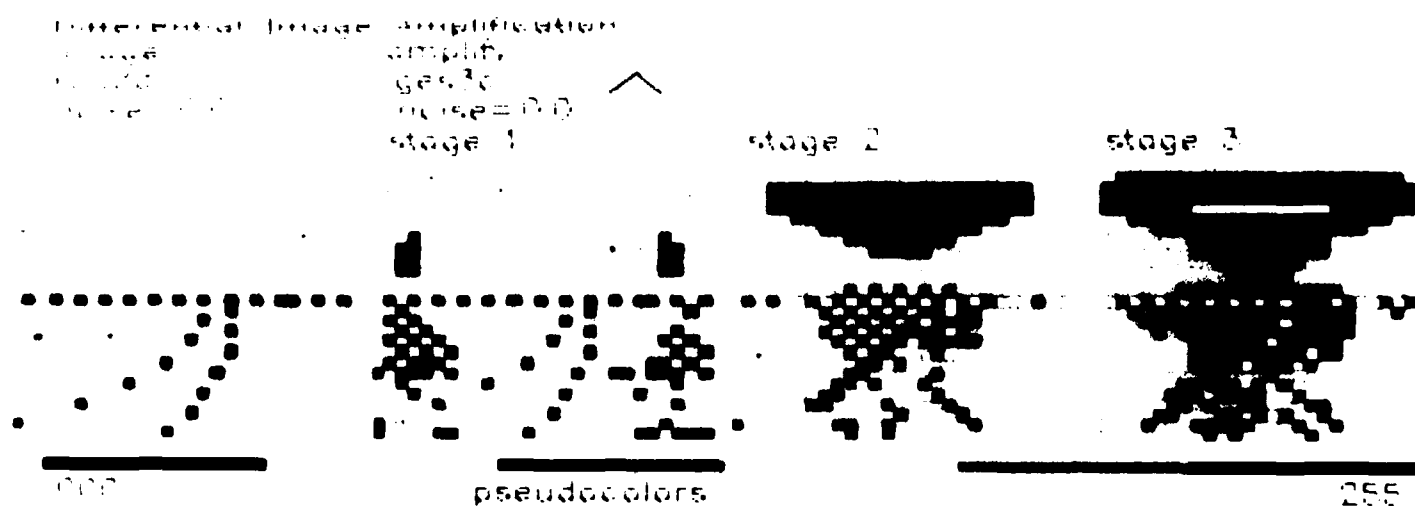


Fig. 19 A skeletal path is added to a skeletal THS and the seek function **ges3c** was used again to modulate ISANS's lateral interactions. Notice that THS and road are "seen".

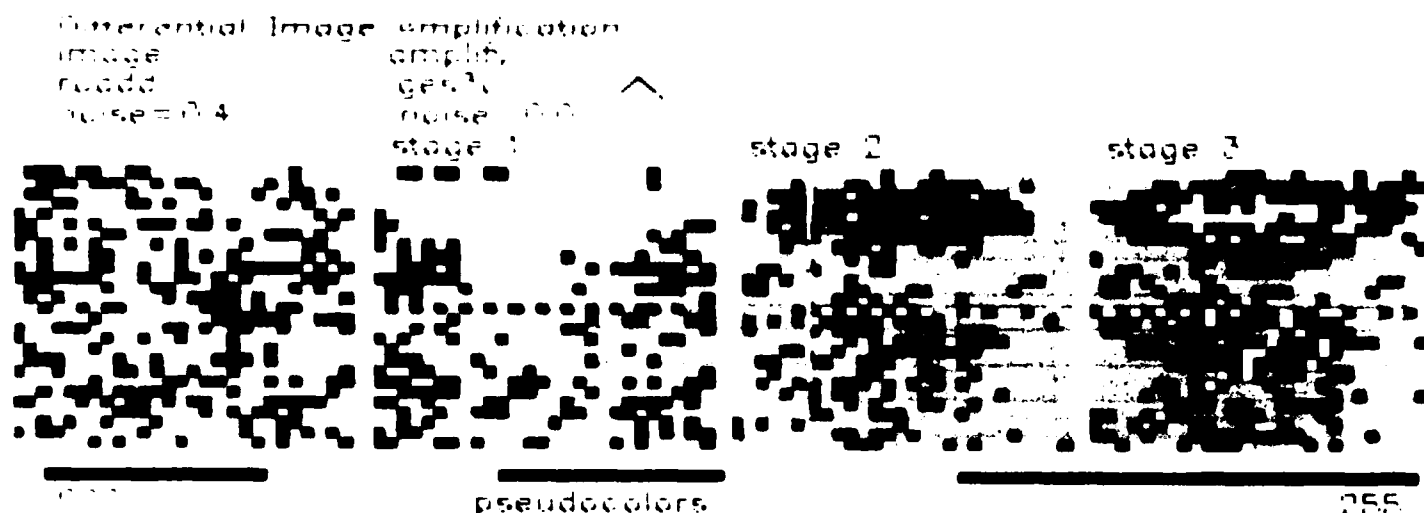


Fig. 20 Same details as Fig. 19 plus noise. Again notice noise suppression and selective amplification of THS and road.

This figure and also figures 13, 14, 15, 16, 17, 18 and 19 strongly suggest that functions such as **ges3c** could be operative in animal vision systems. Furthermore **ges3c** ability to seek such a different variety of THSs demonstrates the feasibility of functions that have holistic goals. In any case perceiving THS, even in heavy noise, requires minimal machinery making it easier to understand how even tiny insects can have astounding navigational capabilities. There also seem to be possible applications for autonomous guidance systems.

END

12-87

DTIC